# Cellular Characterisation of small Open Reading Frame Function in *Drosophila melanogaster*

Unum Amin

Doctor of Philosophy

School of Life Sciences

University of Sussex

Submitted February 2016

# Declaration

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree.

Signed.....

UNIVERSITY OF SUSSEX

Unum Amin

Thesis submitted for the Degree of Doctor of Philosophy

# Cellular Characterisation of small Open Reading Frame function in *Drosophila melanogaster*

## Summary

As our knowledge of the genome expands, so does our understanding of the characteristics of what we define as genes. Small Open Reading Frame (smORF) genes have eluded gene annotation until very recently, and evidence is mounting that these very short nucleotide sequences encode functional peptides that are ≤100 amino acids in size. From work conducted in the fruit fly, our lab has successfully characterised three Drosophila smORFs, of which two have been shown to have a function in higher vertebrates, including humans. The functional characterisation of one of these conserved smORF encoded peptides (SEPs), Hemotin, is presented in this thesis. Though the overall number is still low compared to the abundance of potential smORF-encoding genes in *Drosophila*, the information gathered here allows us to speculate on the wider role of smORF peptides through cell-based imaging studies conducted on *Drosophila* cells. Here, I will discuss the various techniques that can and should be employed in order to study the functions of SEPs. Chapter III describes the various phenotypic studies conducted on the *Hemotin* smORF which is expressed in Drosophila haemocytes, and are integral to the fruit fly immune system. This study showed that connecting subcellular localisation of an SEP to a direct functional assay in cells can reveal functional characteristics of the peptide for further study. Chapter IV details the results from a tagging-transfection assay, which began initially as a way to independently corroborate the translation of smORF mRNAs that were assessed as such by Ribosome Profiling. This experiment resulted in the discovery of several mitochondrial-localised SEPs in *Drosophila* S2 cells, opening the door for the direct functional assay described in Chapter V. The results from a small-scale RNAi screen conducted on the mitochondria of S2 cells provided a reliable read-out for functionality of a large proportion of the smORFs that were screened. This assay can potentially be used as a phenotypic read-out of mitochondrial-SEP function in any cell or tissue type. Elucidation of smORFs and the functions of the peptides that they encode will help us to expand the *Drosophila* proteome, along with providing evidence of their functionality across every organism in which they are found. Considering that characterised SEPs play very important roles in physiology and health, it is time for smORFs to be acknowledged as the important genomic elements that they are.

# Dedication

for Shamim Mazhar (Nano)

# Acknowledgements

I would like to thank my PhD supervisor, Prof. Juan Pablo Couso, for his guidance, support and patience during my time in his Lab. I would also like to thank my co-supervisor Dr. Claudio Alonso for blue-sky meetings and thoughtful insights during my PhD. I am grateful for the support and guidance from Julie Aspden through the latter half of my PhD studies. Thanks for being my teacher and friend, the old school way. I am grateful to Inyaki Pueyo for his immense knowledge and for managing to publish the Hemotin project against-all-odds. Rose Phillips deserves thanks for molecular work and always being ready to go dancing and have a laugh. Sarah Bishop for help with fly work and the fun chats. Thank you to Emile Magny for countless answers to fly genetics related questions and optimistic support through this journey. Thank you to my husband and PhD student in arms, Ali Mumtaz, for putting up with me through my highs and lows during this time, for proofreading the results and indulging my flights of fancy about smORFs and their functions. I would also like to thank former lab members: John Chesebro for for revising chapters of this thesis, as well as proofreading the final version. Miguel Céspedes for showing me the ropes and teaching me how to infect flies. Thank you to Roger Phillips for microscopy advice and helpful talks during my PhD (and MSc). Thank you also to the School of Life Sciences and the University of Sussex for the Graduate Teaching Assistantship, for which I had the opportunity to study here. Finally, I would like to thank my family for all the love, support and visits during my time spent here in Brighton. A very special thanks to my viva voce examiners Prof. Simon Morley and Dr. Stephen Brown.

# Table of Contents

## List of Figures

## List of abbreviations

Small Open Reading Frame (smORF)
Small ORF encoded peptide (SEP)
Upstream Open Reading Frame (uORF)
Deoxyribonucleic acid (DNA)
Ribonucleic acid (RNA)
Nucleotide (nt)
Messenger RNA (mRNA)
Non-coding RNA (ncRNA)
Untranslated Regions (UTR)
Non-coding-RNA ORF (ncr_ORF)
Alternative Open Reading Frame (AltORF)
Complementary DNA (cDNA)
Ribosome bound fragment (RBF)
Model organism Encyclopaedia of DNA Elements (modENCODE)
Reads Per Kilobase per Million (RPKM)
Small-interfering Ribonucleic acid (siRNA)
Transmembrane alpha-helix Hidden-Markov Model (TMHMM)
Mitochondrial Target Peptide (mTP)
Signal Peptide (SP)
RNA interference (RNAi)
Post-transcriptional gene silencing (PTGS)
Third Larval instar stage (L3)
White Pre-pupal stage (WPP)
After pupal formation (APF)
Yellow fluorescent protein (YFP)
Glutathione S-transferase (GST)
Laser Scanning Microscopy (LSM)
Fluorescein isothiocyanate (FITC)
4', 6-diamidino-2-phenylindole (DAPI)
Green Fluorescent Protein (GFP)
Ribosomal protein 49 (RP49)
Paraformaldehyde (PFA)
Distilled H2O (d-H2O)
Methanol (MeOH)
Phosphate Buffered Saline (PBS)
Phosphate Buffered Saline + Triton X (PBT)
*Drosophila* Haemocyte Isolation Medium (DHIM)
Optical density (OD)
Occupied Area Index (OAI)
Double-stranded RNA (dsRNA)
Fetal Bovine Serum (FBS)
Lysogeny broth (LB)
Polymerase Chain Reaction (PCR)
Reverse transcription PCR (RT-PCR)
Ethylenediaminetetraacetic acid (EDTA)
Immunohistochemistry (IHC)
Serum Antibody (SAB)
Immunopurified (IP)
Endoplasmic reticulum (ER)
Central Nervous System (CNS)
Inner mitochondrial membrane (IMM)
Outer mitochondrial membrane (OMM)
Mitochondria associated membranes (MAM)
Mitochondrial DNA (mtDNA)
Adenosine triphosphate (ATP)
Adenosine diphosphate (ADP)
RNA-induced silencing complex (RISC)
Alamar Blue (AB)
*Oregon-R (Or-R)*
Anti-microbial peptides (AMP)
Pattern recognition receptor (PRR)
Microbial-associated molecular patterns (MAMP)

Drosophila inhibitor of apoptosis protein (DIAP1)
Cell division control protein 2 (Cdc2)
Broad-Complex (BR-C)
Ecdysone receptor (EcR)
Stannin (Snn)
Peroxidasin (Pxn)
Serpent (Srp)
Croquemort (Crq)
Hemolectin (Hml)
Vascular endothelial growth factor (VEGF)
VEGF receptor (VEGFR)
Ultraspiracle (Usp)
Lozenge (lz),
Friend-of-GATA (FOG)
U-shaped (Ush)
Phosphatidylserine (PS)
GTP hydrolase enzymes (GTPases)
Mitochondrial open reading frame of the 12S rRNA-c (MOTS-c)
Modulator of retrovirus infection homolog 1 (MRI-2)
Tarsal-less (tal)
Sarcolamban (scl)
Peroxisome proliferator-activated receptor-gamma coactivator 1 (PGC-1)
Sarco-endoplasmic reticulum calcium transport ATPase (SERCA)
Heat shock protein complex 300 (HSPC300)
Ras-related C3 botulinum toxin substrate 1 (Rac1)
Tumor Necrosis Factor (TNF)
Dynamin-related protein 1 (Drp1)

# Chapter I - General Introduction

Our lab has been focused on identifying and characterising smORF genes since the serendipitous discovery of *tarsal-less* (*tal*) (Galindo *et al.* 2007). An open reading frame (ORF) is defined as a potentially protein coding stretch of DNA or RNA that contains a start codon (ATG) and ends in a stop codon (TAG, TAA, TGA). A small-Open Reading Frame (smORF) gene is identified by the presence of an ORF that encodes a peptide which is less than or equal to 100 amino acids in length (Basrai 1997). *tal* is a smORF gene that was found by our lab while studying the developmental patterning of *Drosophila melanogaster* leg morphology (Galindo *et al.* 2000; Galindo *et al.* 2005 Galindo *et al.* 2002; Couso *et al.* 1998) and had previously been classified as a non-coding RNA gene. However, our lab demonstrated that four of the five tandemly arrayed smORFs in the Tal transcript give rise to smORF-encoded peptides (SEPs) of 11, 12 and 32 amino acids in length. These peptides are the shortest peptides ever discovered in metazoans (Galindo *et al.* 2007). Since then, it has been shown that Tal is involved in the differentiation of the second, third and fourth tarsal subsegments in the legs of the fruit fly, and is required for embryonic development (Galindo *et al.* 2007; Pueyo and Couso 2008). This study was one of the first to highlight the fact that SEPs are translated and perform important functions in the fruit fly (Kondo *et al.* 2007; Galindo *et al.* 2007). If such a small peptide could determine vital structures of the adult fly morphology, as well as embryonic viability, it could be possible that there are other previously overlooked small peptides encoded in the *Drosophila* genome.

**The challenge of identifying smORFs**

In order to begin characterising smORFs, we must be able to determine other potential candidates for functional studies. The most logical place to look for further smORF genes to study would be the genome. The challenges of identifying smORFs begin with the first full genome to be sequenced, which was that of the budding yeast (*Saccharomyces cerevisiae)* (Goffeau *et al.* 1996). In the two decades since, mass sequencing projects have enabled a vast amount of genomic data to be generated. These data have revealed that there is a huge amount of information encoded within genomes,

and led to the subsequent annotation of thousands of genes. These annotation approaches struggle to cope with the sheer volume of data that has been generated, and it is only natural that there is some exclusivity generated by the annotation exercises. In the *S. cerevisiae* genome, hundreds of thousands of ORFs were identified across a large size range, encoding for peptides across a range of hundreds to thousands of amino acids. During this time, the presence of a large number of putative smORFs was shown as an outlier to this distribution, as they amounted to a total of ~260,000 ORFs that were 2-99 codons in length (Basrai *et al.* 1997). Of these hundreds of thousands of smORFs, only four hundred showed clear homology to known proteins and almost one hundred functional SEPs had already been verified in yeast through genetic and biochemical experiments (Basrai *et al.* 1997). This highlighted the importance of retaining some, if not all smORF encoding genes in genome annotations. However, the remaining ~259,000 smORFs were initially left out of *S. cerevisiae* genome annotation by imposing a minimum 300 nucleotide (nt) ORF length cut-off (Fickett 1995) due to their sheer abundance (Fickett 1995; Das *et al.* 1997). Despite the discovery of smORFs and evidence of functionality of peptide products, the smORF genes being largely excluded from yeast genome annotations set a precedent for subsequent genome annotation exercises in other organisms.

When smORF discovery is approached at a genome-wide scale in *Drosophila*, smORFs fail to obtain high conservation scores in relation to controls designed for canonical genes, and must be treated differently due to their small size (Ladoukakis *et al.* 2011). Most systematic bioinformatic searches have relied on evolutionary conservation of DNA sequences as a criterion to identify novel genes, which is based upon the theory that conservation of gene sequences implies functionality (Das *et al.* 1997; Clark *et al.* 2003). Searches for novel protein coding genes have also been targeted to highly transcribed (euchromatic) regions of the genome, based on the assumption that essential genes must be undergoing active transcription (Misra *et al.* 2002; Consortium 2004; Stein 2004). Typically, the presence of a start and stop codon and some sort of sequence conservation and homology would be enough to predict the coding potential of an ORF. However, analyses of the human and mouse transcriptomes have shown that large proportions of the mammalian genomes is transcribed into RNA without any further

translation of the RNA into a peptide (Dinger *et al.* 2008). Experimental evidence also exists for conserved developmental genes that have been found to employ the RNA rather than the peptide as the functional product of the gene. (Pollard *et al.* 2006). The transcripts from these kinds of genes have generally been classified as non-protein coding, or non-coding RNAs (ncRNAs).

Potentially, twenty-six percent of human and thirteen percent of mouse RNA is polyadenylated ncRNA (Okazaki *et al.* 2002; Ota *et al.* 2004). This means that polyadenylation cannot be used as an exclusive marker to identify protein-coding transcripts. ncRNA transcripts can be classified into short or long ncRNA. Short ncRNAs include microRNA (miRNA) and small-interfering RNA (siRNA), which have been increasingly shown to be involved in the regulation of gene expression (Aboobaker *et al.* 2005; Nguyen *et al.* 2006; Haley *et al.* 2008). The very abundant 'long' non-coding RNAs (lncRNA) transcripts undergo post-transcriptional processing such as polyadenylation and splicing, and fall in the same size range as smORF transcripts. Therefore, due to the 300 nucleotide minimum ORF length cut-off discussed previously, many smORF-containing transcripts are annotated as putative non-coding RNAs as there is no way to differentiate between them.

In order to identify other smORFs in *D. melanogaster*, our lab conducted a systematic bioinformatics search for putative smORFs in traditionally non-coding regions of the *Drosophila* genome. The bioinformatics pipeline included filters for size (ORFs coding <100aa), the presence of both start (ATG) and stop codons and conservation between *D. melanogaster* and *D. pseudoobscura* (divergence of 25-55mya) (Ladoukakis *et al.* 2011) to identify 4,561 putative smORF candidates. This study used two further evolutionary conservation criteria: 1) the ratio of synonymous to non-synonymous amino acid substitutions of the codons in the smORF and 2) the conservation of sequences around the smORF in syntenic regions of the *D. pseudoobscura* and *D. melanogaster* genomes, to narrow this number down to 401 'high confidence' smORFs. Together, this data would imply that these identified smORFs have functional, protein-coding potential.

By combining information obtained from other functional studies and bioinformatics predictions, this study places the potential number of *D. melanogaster* smORFs within the range of predicted smORFs found by similar studies in *S. cerevisiae*, *M. musculus*,

and *A. thaliana* (Kastenmayer *et al.* 2006; Frith *et al.* 2006; Hanada *et al.* 2007; Ladoukakis *et al.* 2011). smORFs have also been found in rice (*Oryza sativa*) by comparing the genome to that of *A. thaliana* and using smORF nucleotide sequence conservation between the diverged plant species to predict that they are 'real' smORFs (Okamoto *et al.* 2014). Analysis of conserved amino-acid sequences (as opposed to nucleotide sequences) resulted in the identification of almost 2000 smORFs in the untranslated regions (UTR) of both canonical mRNAs and within non-coding RNAs in humans, mice, zebrafish, fruit flies and the nematode *C. elegans* (Mackowiak *et al.* 2015).

Of these bioinformatically predicted smORFs, published ribosome profiling data predicts putative translation of more than 100 novel smORFs, and mass spectrometry data further corroborates 70 novel SEPs. This leads to the conclusion that a significant proportion of smORFs may be translated and encode functional peptides, with evidence for their putative functions demonstrated by their conservation across phyla and between species. The 401 smORFs identified by our bioinformatics study are very strong candidates for functionality as they have evidence of transcription and conservation (Ladoukakis *et al.* 2011), but so far only two of these smORFs have been proven to be translated into SEPs and are functional peptides in *Drosophila* (Magny *et al.* 2013; Pueyo *et al.* 2016, in print). The general lack of peptide level evidence and homology to known protein coding genes has compounded the ambiguity around the question of whether or not smORFs are translated into functional peptides. It is clear that without proof of translation, we cannot determine smORF function, therefore other ways have been sought to circumvent the issue of bioinformatically predicting the coding potential of smORFs.

**Assessing smORF translation**

Of the 21,780 total annotated protein coding ORFs in the *D. melanogaster* genome, 829 are coding for peptides <100aa (3.8%) (Flybase (FB), http://flybase.org/), which is similar to the total proportion of annotated smORFs in zebrafish (2%), mice (2.2%) and humans (1.9%) (Aspden *et al.* 2014). Even though these smORF genes are annotated as putatively protein coding, there is very little experimental evidence of their translation,

therefore these can fluctuate between protein-coding and non-coding between different releases of the genome. Thus in order to study them any further; it becomes essential to somehow assess the translation of these smORFs, especially in the absence of a detected peptide. Standard Mass Spectrometry techniques have generally not been successful at the detection of SEPs and genome-wide studies have only revealed the translation of a few SEPs in *Drosophila* (Brunner *et al.* 2007).

Ribosome profiling involves the deep sequencing of ribosome footprints, which are fragments of mRNA that are protected from nuclease digestion by binding of ribosomes (Ingolia *et al.* 2009). These sequenced Ribosome-bound mRNA fragments (RBFs) are mapped to the transcriptome to highlight transcripts that are undergoing translation *in vivo*. This technique has allowed the identification of which regions within the mRNA are being translated and also reveals the abundance of these translated regions, which can be an indicator of cellular or tissue requirement of the product. Several ribosome profiling studies performed by other research groups have demonstrated the translation of a variety of novel ORFs in various organisms (Ingolia *et al.* 2009; Vanderperre *et al.* 2013; Duncan *et al.* 2014; Aspden *et al.* 2014; Bazzini *et al.* 2014). These novel ORFs occur outside of annotated protein coding regions, such as in 5' (upstream-ORFs) and 3' untranslated regions as well as in non-coding RNAs (ncr_ORFs) and alternative reading frames of known protein coding sequences (AltORFs) and have been reviewed in (Ingolia *et al.* 2014).

By using an assessment of ribosome footprint periodicity through ribosome profiling methods, Bazzini *et al.* (2014) have identified several hundred translated smORFs in both zebrafish and human cell lines. Evolutionary conservation of these novel smORFs between organisms helps to corroborate these results (Vanderperre *et al.* 2013), especially for those smORFs which are believed to be important regulators of development and physiology in higher vertebrates (Bazzini *et al.* 2014). Ribosome profiling studies in yeast show that 24% of previously considered ncRNAs could potentially be undergoing translation (Duncan *et al.* 2014). This study also showed the translation of almost two thousand upstream-ORFs (uORFs), which are smORFs that occur in the 5'UTR of a longer, canonical transcript. Some uORFs have been

experimentally shown to affect the translational regulation of the longer downstream ORFs (Zhang *et al.* 2005).

Using the relatively new technique of ribosome profiling, our lab attempted a genome-wide assessment of smORF translation in order to determine how many and which *D. melanogaster* smORFs are translated. For our Ribo-Seq experiment, the technique was applied to S2 cells, which are derived from a spontaneously immortalised embryonic cell line obtained from late-stage *D. melanogaster* embryos (Schneider 1972). smORF coding sequences are by definition shorter than those of canonical protein coding genes, hence they present a smaller target for ribosome binding. This means there are relatively fewer RBFs that can arise from a smORF mRNA in the sequencing data, which creates a challenge for classifying smORFs as translated when compared to the number of footprints derived from longer transcripts. This experimental limitation was overcome by developing a modified version of ribosomal profiling called Polysomal Ribosome Profiling (Poly-Ribo-Seq)., which incorporates polysome fractionation into the initial stages of the technique. By limiting the starting material to transcripts bound by 2-6 ribosomes, we hoped to increase the proportion of RBFs in the library arising from smORF transcripts. The Poly-Ribo-Seq experiments revealed the translation of ~230 annotated smORFs in *Drosophila* S2 cells (Aspden *et al.* 2014). Our Ribosome profiling study also revealed the translation of 2,708 uORFs and 313 smORFs in non-coding RNAs in *Drosophila* S2 cells (Aspden 2014).

Ribosomal profiling is gradually becoming an accepted high-throughput assessment of genome-wide translation for annotated protein coding regions but significant debate remains with regards to the real translation and function of these novel ORFs. It has been argued that even though RBFs align to many smORF sequences within various transcripts, this association does not constitute proof of translation of these smORFs, until a peptide product is detected (Chew *et al.* 2013). A number of bioinformatic approaches have been developed to try and identify translated ORFs by using confidence scoring, but these studies have their own challenges (reviewed in Mumtaz and Couso 2015). Therefore, until a much larger proportion of smORFs have proof of translation at the peptide level, there still exists a need to corroborate some, if not all, of the SEPs from smORFs discovered by ribosome profiling. However, few techniques

match the scope and sensitivity of ribosome profiling. Mass spectrometry conducted on a small protein fraction from S2 cells could only corroborate about 25% of the total annotated smORFs that are found to be translated by Poly-Ribo-Seq (Aspden *et al.* 2014).

Some research groups are choosing to approach characterisation of smORFs by detecting SEPs directly through Mass Spectrometry and one such study resulted in the discovery of 90 SEPs, of which the majority are encoded by novel, previously uncharacterised smORFs in human cells (Slavoff *et al.* 2012). As the pool of corroborative information grows, so does the confidence in the data obtained from Ribosome Profiling experiments. This has resulted in the general consensus that smORFs are translated, despite the low rates of SEP detection and lack of functional characterisation (Ingolia 2014; Andrews *et al.* 2014; Mumtaz and Couso 2015; Brar *et al.* 2015;). However, there are still some doubts and it could be considered that the low numbers of SEP-encoding ORFs are actually contaminants in the larger pool of genuine ncRNAs (Chew *et al.* 2013). Therefore further work must be done to differentiate the coding-features of smORFs to support their existence as proteins. One of the ways to do this is to prove that SEPs are functional at the peptide level.

**The functions of smORF-encoded peptides**

There is significant evidence that some smORFs can have homologues across vast evolutionary distances (Ladoukakis *et al.* 2011; Slavoff *et al.* 2012; Aspden *et al.* 2014; Bazzini *et al.* 2014). This means that the functional characterisation of more SEPs could open the door for hundreds of previously uncharacterised peptides across animal phyla. Known SEP functions range across various biological processes, such as development, regulation of metabolism, transcriptional regulation and ion transport across membranes. These type of functions are not distinct to SEPs, or exclusive from the function and regulation of longer, canonical proteins. There are a few smORFs that have been functionally characterised in humans, such as the smORF found in mitochondrial 16S RNA, which produces a 24aa SEP called Humanin. Expression of Humanin protects cells from amyloid-beta mediated apoptosis and inhibits proteins of the pro-apoptotic cascade (Lee *et al.* 2013). Another human smORF encodes the 16aa MOTS-c

SEP, which is localised to the mitochondria of muscles and regulates cellular metabolism. It inhibits the folate cycle and increases the sensitivity of muscles to insulin, preventing weight gain and metabolic disease in mice (Lee *et al.* 2015). MRI-2 is a 69aa long SEP discovered in human cells, which mediates DNA double-strand break ligation by non-homologous end joining, therefore playing a role in DNA repair (Slavoff *et al.* 2014).

In *Drosophila,* The 65 amino acid long SEP Reaper is a regulator of the apoptotic pathway (White *et al.* 1994). This is perhaps one of the most important SEPs that has been found in *Drosophila*, as the function of Reaper sits at the convergence point of the different signals that lead to the induction of the cell death program. Reaper has also been shown to induce apoptosis in *Xenopus* egg extracts, showing that it can perform its biological function in vertebrates systems as well (Evans *et al.* 1997). This is one example of an unavoidable smORF discovery, leading from the investigation of a major cellular process, but other smORFs may not be playing such an overt role. Other previously characterised *Drosophila* smORFs show a broad range of functions that could be very important for fly life and health. Spargel is a 71aa SEP which was discovered in *Drosophila* S2 cells upon treatment with the insulin-regulated transcription factor dFoxo (Tiefenböck *et al.* 2010). Spargel functions in parallel to this transcription factor to help fine-tune insulin signalling. Spargel is a novel homolog of mammalian peroxisome proliferator coactivator-1 (PGC-1) (Gershman *et al.* 2007), which has been found to control mitochondrial gene expression in mammals. In *Drosophila,* over expression of Spargel leads to increased mitochondrial activity which extends the adult fly life-span (Rera *et al.* 2011); thus Spargel plays a significant role in the integration of mitochondrial activity, gene expression and metabolism. Another *Drosophila* SEP, HSPC300 (75aa), acts as a regulator of actin in the WAVE/SCAR complex, which is involved in the downstream effects of the cytoskeletal Rac1 GTPase signal transduction pathways (Qurashi *et al.* 2007). These signals dictate cellular movements through the regulation of actin and HSPC300 is particularly important in the nervous system, where it helps to control embryonic CNS axon and synaptic morphologies via these pathways.

The 11aa Tarsal-less peptide can perform the entire range of the functions of this polcistronic smORF gene as long as the peptide contains the LDPTGXY motif, which is phylogenetically conserved across insects, indicating its importance (Kondo *et al.* 2007; Galindo *et al.* 2007). In addition to Tal, which functions as a cell signal, our lab has also characterised the function of *sarcolamban (scl)*, which encodes two SEPs of 28 and 29 amino acids in length. Scl was shown to be a homologue of the vertebrate Sarcolipin and Phospholamban peptides (Magny *et al.* 2013). These peptides are structurally and functionally shown to be playing a role in the regulation of the $Ca^{2+}$ SERCA pump in both vertebrates and flies. The SERCA pump is found in the sarcoplasmic reticulum of heart muscles in most mammals and also in mouse skeletal muscle (Anderson *et al.* 2015). Changes in the expression of *sarcolamban* and its homologues have an effect on the rhythmic beating of the heart, which lowers the overall health of the fly (Magny *et al.* 2013). The functional homology of Scl in higher eukaryotes, as well as of some of the other SEP examples given here, can be considered further evidence for why the study of smORFs in *Drosophila* is a subject of wider relevance, not just for *Drosophila* development and physiology, but for vertebrates as well.

Approaching the question of smORF function using the fruit fly as a model has uncovered the vital functions of some novel smORF encoded peptides. However, research on SEP functions has been serendipitous, and follows smORF gene discovery. This illustrates a need for a more targeted approach in the detection of SEPs and their functions. Studying SEP function at the organism-level is vital to giving relevance to the wider function of smORFs, and for this purpose, *Drosophila* is probably one of the best eukaryote models for the functional characterisation of these novel genes, because of its ease of genetic manipulation and extremely well annotated genome. However, the methods employed by these single gene studies cannot currently be extended to a genome-wide smORF functional study due to the various challenges of assessing very small proteins. Therefore, each smORF that has been characterised so far, requires years of research on a particular pathway or system while breaking new ground with a range of diverse techniques. If it becomes possible to unravel the functions of smORFs in a high-throughput manner in *Drosophila*, this would open the door for similar studies in vertebrates, where time and experimental limitations are even bigger concerns.

The only genome-wide study conducted specifically on smORF function, has been conducted in *S. cerevisiae,* in which less than 10% of the smORFs tested led to a mutant phenotype in relation to the growth of yeast cells (Kastenmayer *et al.* 2006). This could be because not all smORFs in yeast are involved in growth, but also because growth can be attributed to several processes occurring in the cell. This highlights the need for establishing a higher-throughput approach to characterise smORFs by gathering more data on specific functions in which SEPs are directly involved, in order to characterise their functions. Genome-wide Ribosome Profiling and Mass Spectrometry experiments have provided us with data about which smORFs are being translated, but only a small minority of these SEPs have any specific localisation data or experimentally corroborated functions. The direct detection of SEPs can be considered vital to understanding smORFs on a broader scale, as they are likely to function where the are found. The ability to label and knock-down smORF gene products would aid in understanding their functional phenotypes.

The main question addressed in this thesis is how to assess the function of smORFs and SEPs in a high or medium-throughput manner in *Drosophila*. This question has been addressed by approaching smORF functions at the cellular level, which provide an experimental system that can be manipulated at a more rapid and wider scale than studies conducted *in vivo*. Most important mechanisms of cellular function remain unchanged across eukaryotic cells, providing a broader scope for this study in higher eukaryotes as well, especially for SEPs conserved in vertebrates, therefore in this thesis, I use *Drosophila* haemocytes and S2 cells to study the localisation and function of smORF-encoded peptides. Like several other *Drosophila* embryonic cell lines, S2 cells display gene expression similar to that of haemocytes (Cherbas *et al.* 2011) and some cell lines also display haemocyte-like motility (Baum and Cherbas 2008). S2 cells are also phagocytic, which is a characteristic behaviour of macrophage-like cells (Elwell and Engel 2005). After pursuing an in-depth study on a single smORF expressed in haemocytes *in vivo* (*Hemotin;* Chapter III). The results from this study were used as groundwork to create a pipeline for the cell based assessment of SEP functions in S2 cells. This pipeline allows us to provide novel insight into the characterisation of several SEPs, which localise to and function within the mitochondria (Chapters IV and V).

# Chapter II - Materials and Methods

## A. Animal rearing

All experimental fly stocks were kept in vials or bottles at 25ºC on a standard cornmeal diet.

## B. Embryo collection

Flies expressing GFP in haemocytes: w; *srpGAL4-UASGFP; crqGAL4-UASGFP* were crossed with *Hemotin* null flies (*w;;Df A4/TM3-YFP*) to establish the stable line: *w; srpGAL4-UASGFP; Df A4/Tm3-YFP.* The TM3-YFP chromosome III balancer was used to maintain this line, as male *Df A4* flies are sterile. The embryos and larvae could be differentiated from *Df A4* homozygous flies due to the presence of the strong YFP in the presumptive eye region. These flies were transferred from vials into egg-laying 'cages' supplemented with apple juice agar plates with yeast paste to collect embryos. After discarding embryos from the first 24 hours, plates were changed at 12-hour intervals, resulting in the collection of a mixed population of embryos between Stage 1 to Stage 14 of embryonic development (Hartenstein 1985). Embryos were collected from the apple juice agar plate with a paintbrush into a basket and washed using distilled-$H_2O$ ($dH_2O$) to prepare for dechorionation. Embryos were dechorionated by adding a few drops of bleach to the collection basket and rocked for ~2-3 min, until the chorion dissolved and embryos appeared shiny; the bleach was washed out using $dH_2O$ and the embryos were blotted to dry.

## C. Embryo fixation and removal of vitelline membrane

Paraformaldehyde (PFA) was used as a fixative to crosslink molecules of the embryo *in situ*. One millilitre of 4% PFA and 1ml of Heptane were added to a 2ml Eppendorf tube. The dechorionated embryos were added to the vial using a paintbrush.The Eppendorf was taped to an agitator at the lowest settings for 20 minutes to allow fixation to occur. After this time, the bottom PFA phase was removed, taking care to ensure that embryos remain in the Eppendorf. One millilitre of 100% methanol (MeOH) was added to the tube and the vials were shook hard for several minutes, until almost all the embryos

sunk to the bottom of the tube, indicative of vitelline membrane removal. The heptane was aspirated from the sample, and after 2–3 washes with methanol, the embryos were stored at –20ºC in fresh 100% MeOH.

**D. Embryo immunohistochemistry**

The Hemotin antibody (Eurogentec) was raised against the N-terminal peptide sequence of Hemotin: RISITPGCSMVATTKLTHSRNSVDIY. Fixed embryos were washed in 1X Phosphate Buffered Saline (PBS), and then 1X PBT (1X PBS + 0.1% Triton X-100). Embryos were then prepared for antibody staining by blocking in 1X PBT-BSA (0.3% Triton X-100, 0.2% Bovine Serum Albumin). Rabbit anti-Hemotin Serum and Immunopurified antibodies were diluted to 1:200, along with 1:500 mouse anti-GFP (Roche) in PBT and incubated overnight overnight at 4ºC. The next day, embryos were incubated for two hours at room temperature with anti-mouse FITC and anti-rabbit Biotin (each at 1:200). Embryos were then mounted on glass slides in Vectashield (Vector). Fluorescent antibody stained embryos were analysed and images were captured using a Zeiss LSM510 Meta point scanning confocal microscope and the Zeiss LSM software (Sussex Centre for Advanced Microscopy, University of Sussex; Brighton, UK). LSM Image Browser and ImageJ softwares were used for image processing.

**E. Primary haemocyte immunohistochemistry**

*Hemolectin_GAL4 (Hml_GAL4)* (Goto 2003) flies were crossed to the *UAS-Hemotin-GFP* fly line. The *w;;Df A4/Tm3-YFP* balanced stocks were used to obtain Hemotin null animals, which do not have the Tm3-YFP marker. *Oregon-R* flies were used as wild type controls. From the tubes of these flies, 6 animals at wandering third instar larval stage (L3) or white prepupal (WPP) stages were collected for each genotype, washed with $dH_2O$ and blotted to dry. Animals were sorted on a fluorescent dissection microscope to differentiate the *Df A4* L3 and WPP. The animals were bled using a sterile needle and forceps in 200µl DHIM (25% FBS v/v Schneider's media: *Drosophila* Haemocyte Isolation Medium (Sampson and Williams 2012)), which consists of filtered Schneider's *Drosophila* medium supplemented with 25% Fetal Bovine Serum (FBS).

The staining protocol is adapted from Zettervall et al. (Zettervall *et al.* 2004), in which haemocytes are fixed and stained *ex vivo.* After removal of the carcasses, the remaining 200µl was transferred to collagen IV-coated coverslips in circular culture dishes to settle for 30 minutes before further staining. After this stage, in the case of organelle dye Lysotracker-RED DND-99 (Life Technologies), the dye was diluted to 1:1000 in DHIM and 200µl of this was incubated with the cells for 15 minutes after removal of plating media. As further fixation would remove the Lyostracker staining, these samples were imaged live Imaging was conducted on a Zeiss LSM 510 confocal microscope with automated prior stage. LSM Image Browser, ImageJ v1.46r softwares were used for image processing.

If samples were to be used for antibody staining, plating media was removed and 200µl of 4% PFA fixative was added to the culture dish for 10 minutes and then the cells were washed with 1X PBT. After this, cells were blocked with 1X PBT-BSA for 1 hour. The primary antibodies used for immunohistochemistry are: rabbit anti-Hemotin (1:200), rabbit anti-KDEL (DSHB) (1:50), mouse anti-GRP75 (Santa Cruz) (1:200), rabbit anti-14-3-3 pan (Santa Cruz) (1:400), mouse anti-beta tubulin (DSHB) (1:500). Cells were incubated with primary antibodies either overnight at 4ºC or for 2 hours at room temperature. After washing with 1X PBT, samples were incubated with anti-rabbit Rhodamine (1:200), anti-mouse Biotin (1:200) or anti-rabbit Biotin (1:200) from all from Jackson laboratories. In the case of Hemotin staining, with biotin labelling followed by Avidin Rhodamine (Jackson) tertiary antibody (1:1000). For F-actin staining, 1:500 Phalloidin-FITC (Life Technologies) dye was added after antibody incubations for 30 minutes at room temperature. After 1X PBT washes, Hoechst nuclear stain (Sigma) was used at a concentration of 1:500 for 10 minutes. Samples were then washed with 1X PBS, covered with Vectashield and stored at 4ºC. Imaging was conducted on a Zeiss LSM 510 confocal microscope with automated prior stage. LSM Image Browser, ImageJ v1.46r softwares were used for image processing. The ImageJ plugin ‘Manders Coefficients’ was used to calculate correlation coefficients of Hemotin-GFP to signal from the various organelle markers.

**F. Mander's Coefficients for colocalisation of markers**

Mander's coefficient is used to quantify the amount of overlap between pixels of different channels that detect signal from different fluorescent markers (Dunn *et al.* 2011). All sample images were quantified by cropping the cells from the main image in order to isolate individual scores for each cell. The R-value for each measurement of colocalisation was then averaged for all the cells of each sample imaged.

**G. Fly Survival Assay**

The fly viability assay was conducted on male and female virgins 0–2 days old from the following lines: *Hemotin* mutant (*Df A4*), wild type *Oregon-R (Or-R)* and positive control *fob1* allele flies (Akber *et al.* 2011). In order to verify the specificity of viability to the *Hemotin* gene, a fly line that rescues the deletion of CG7691 in a *Df A4* background (made by J.I. Pueyo) was used. The collected flies were kept at a maximum density of 20 flies per vial. For each strain, we kept three classes: untreated, wounded and infected. Bacteria were grown from stock cultures stored at –80ºC. A toothpick amount of each bacterial strain was placed in culture flasks containing 10ml of Lysogeny Broth (LB) medium and cultures were grown overnight at 37ºC. The next day, 1ml of the starter culture was placed in Erlenmeyer flasks containing 50ml of LB media and allowed to grow at 37ºC; bacterial concentrations were checked every hour after an initial 2-hour growth phase. The Gram-positive, pathogenic bacteria *Micrococcus luteus*, and the Gram-negative pathogenic bacteria *Enterobacter cloacae* were used at a final concentration of optical density (OD) 0.75 and 0.84, respectively, in LB medium. The non-pathogenic DH5α *Escherichia coli* was grown to an OD of 0.95 and then spun down in a cooled centrifuge (5000g) to pellet the bacteria. The pellet was then resuspended in 0.5ml of PBS for a final concentration of approximately OD 0.95, which is theoretically impossible, but in relative terms, this would be a very high concentration of DH5α *E. coli.* Fly infections were conducted by pricking the flies in the ventrolateral abdomen, under the haltere, with a sterile 0.25 gauge tungsten needle dipped into the bacterial cultures. As a control, flies of each strain were wounded in the abdomen with a tungsten needle dipped in sterile 1X PBS. The flies that remained alive were then counted over the next 60 days.

**H. *In vivo* imaging of embryonic haemocytes**

*w; srpGAL4-UASGFP; Df A4/Tm3-YFP* heterozygous flies were allowed to lay embryos on apple juice agar plates. These embryos laid by these flies were sorted for homozygous and heterozygous conditions based upon the presence of the TM3-YFP marker. After dechorionation, embryos were mounted on double-sided tape stuck to a glass slide, with the ventral side towards the coverslip and covered in 10S Voltalef Oil (SIGMA); a coverslip was applied to the slide and affixed using nail varnish. This is similar to the protocol used by the Wood Lab (University of Bath) (Evans 2010) (Evans 2010a). Imaging was conducted on the Zeiss LSM 510 confocal microscope. *In vivo* time lapses were obtained as well as stills, using a Z-stack taken every 2–4 minutes for 30 min to 1.5 h. Image analysis was conducted with Freeware software, ImageJ. The ImageJ Plug-in 'Grouped ZProjector' was used to make projections of each frame.

**I. *In vivo* imaging of pupal haemocytes**

*w; srpGAL4-UASGFP; Df A4/Tm3-YFP* flies were allowed to lay eggs in tubes and adults were removed frequently to prevent crowding. Brown pupae (40–50h after puparium formation) were recovered and sorted for homozygous and heterozygous conditions. The selected pupae were washed with $dH_2O$ and blotted dry. Pupae were mounted ventral-side down on double-sided tape applied to a glass slide. The pupal cases were dissected in a manner similar to that of Moreira *et al.* (2011) (Moreira *et al.* 2011), by removing a window in the pupal case over the thorax. After application of 10S Voltalef Oil on the thorax region, a ring of petroleum jelly was made around the samples, a coverslip was rested on this ring and pressed down on the sample. The slide was raised on the microscope stage by adding additional pieces of glass slide to the edges to allow deeper focal length. Imaging of live haemocytes was conducted on an inverted Zeiss LSM 510 with automated prior stage, using a Hamamatsu ORCA-ER C4742-95 digital camera. A Z-stack was taken every 2–4 minutes for 30 min to 1.5 hours. Image analysis was conducted with ImageJ.

**J. Measuring Occupied Area Index of haemocyte vacuoles**

Cell and vacuole sizes were measured using ImageJ. Lines were drawn around GFP-labeled haemocytes and the dark vacuoles in the cytoplasm with the drawing tool used

to label ROIs and then measured using the Measure tool to show the area of the ROI of the cells and vacuolated areas. The sum of the vacuolated areas of each cell was divided by the overall cell area to calculate the vacuolar Occupied Area Index (OAI) for each cell, and averaged.

**K. Primary haemocyte culture *ex vivo***

Flies were allowed to grow in stock tubes; L3 and WPP were collected and bled with a 0.25mm gauge needle in DHIM. These were plated on collagen-coated coverslips and allowed to adhere for 30 minutes. *Ex vivo* imaging and image analysis was conducted using an inverted Zeiss Axiovert 200M microscope with automated prior stage with a Hamamatsu ORCA-ER C4742-95 camera. *Ex vivo* time lapses were 20 minutes at 15 seconds/frame, as conducted by Sampson *et al.* (Sampson *et al.* 2013). All *ex vivo* experiments consisted of haemocytes excised and isolated from a minimum of 6 L3 or WPP, which was repeated a minimum of two to three times per life stage/experimental group. ImageJ v1.46r was used to make time-lapse movies for subsequent analysis of morpho-behavioral haemocyte types.

**L. S2 cell culture, transfection and imaging**

S2 cells were cultured in 10% FBS (v/v) Schneider's media. The cells were grown to a confluent stage (2 days after splitting) and then resuspended in fresh media and diluted at lower dilutions in new flasks. For imaging and transfections, 1ml of S2 cells at a concentration of 250,000 cells/mL were plated on acid-treated coverslips in 6-well plates. The smORF transfections were conducted by combining 1µg of plasmid DNA of each transfection in order to have consistent levels of DNA in the S2 cells. The smORF fusion constructs were combined with a lipid-based transfection reagent, Xtreme Gene HP (Roche). The plasmid and transfection reagent mix was added to the cells and then left for 48 hours in order to allow the plasmid to be expressed by the host cells translation machinery. The cells were then fixed for 20 min with 4% PFA, washed with 1X PBS and 1X PBT, then blocked with PBT + 2% BSA (wt/vol), before IHC with primary mouse anti-FLAG M2 antibody (1:1000; Sigma) and secondary anti-mouse FITC (1:400; Jackson). For subcellular localisation experiments, cells were incubated

for 30 min with 1:500 Rhodamine-Phalloidin (Life Technologies) to highlight F-actin. All transfections were incubated for 10 min with Hoechst (Sigma) according to manufacturer's instructions for nuclei staining and mounted with Vectashield (Vector Labs). For Mitotracker experiments, cells were incubated in 500 nM Mitotracker Red CMXRos (Life Technologies) for 45 min. Cells were then fixed and the immunohistochemistry protocol for anti-FLAG was conducted as above, followed by DAPI to stain the nuclei. Imaging was conducted using a Zeiss 63X Plan Apochromat Oil Immersion lens on the LSM510 Axioskop 2. For colocalisation analysis Z-stack images were taken with slice interval of 0.15 μm, and the ImageJ plugin 'Mander's Coefficients' was used to calculate correlation coefficient of FLAG to Mitotracker signal with at least 15 cells per replicate, with three replicate transfections.

**M. Alamar Blue Assay**

Confluent S2 cells were resuspended in Serum-Free Schneider's media and counted on the haemocytometer, after which they were diluted by volume to required densities. The dsRNAs used in this assay were made at the RNAi screening facility at the University of Sheffield (Sheffield, UK). One micro litre of 0.1μg/μl dsRNA was added to the bottom of the wells and 100μl of cells were quickly added. One hour after the addition of cells, the wells were supplemented with 20% FBS Schneider's media for the normal amount of FBS used to culture S2 cells. The plates were sealed with ParaFilm and incubated at 25ºC for 48-96 hours, depending on the experiment. The growth media was then aspirated from the wells and 200uL of fresh Serum-Free media containing 10% Alamar Blue (ThermoFisher) was added to each well. The plates were incubated at 25ºC for up to 6 hours and triple fluorescent readings were obtained for the samples at hourly intervals in the GloMax Plate Reader (Promega)

**N. Verification of positive control dsRNAs by RT-PCR**

Total RNA was extracted from S2 cells transfected with dsRNA against the controls DIAP1, cdc2 and BlueScript in the following way. $2x10^6$ cells were plated in 1ml of Serum-Free media in each sample well to which 8μg of dsRNA was added; after 1 hour, 1ml of 20% FBS Schneider's media was added. The plate was sealed and incubated for

48 hours at 25ºC. After this time, cells were resuspended in the growth media and transferred to 2ml Eppendorf tubes. The samples were spun at 5000g for 10 minutes in a chilled centrifuge (4ºC) to pellet the cells. After pelleting, the media was removed and cells were washed with ice-cold 1X PBS and spun down twice. After removal of PBS, 800µl of the TRIzol reagent (Invitrogen), which contains a mixture of phenol, guanidium thiocynate, and ammonium thiocynate, was added to the cell pellet. The samples were homogenised by pipetting and stored at -80ºC. After the addition of chloroform:isoamyl (24:1) the tubes were incubated for 5 minutes at room temperature, spun to separate phases and the resulting supernatant was transferred to a sterile, nuclease-free tube.

In subsequent steps, the RNA was precipitated with isopropanol and washed with ethanol before being resuspended in nuclease-free $H_2O$ and quantified using a NanoDrop Spectrophotometer. cDNA was synthesised using the RETROscript Reverse Transcription Kit (Ambion), following the protocol for the 'Two-step RT-PCR with heat denaturation of RNA' procedure provided by the manufacturer. One microgram total RNA was combined with either random decamers or oligo(dT) primers and nuclease-free water, then denatured at 80ºC before the addition of the remaining RT reagents: 10X RT buffer, dNTPs, RNase inhibitor, and the M-MLV Reverse Transcriptase. Reverse transcription of cDNA was done at 42ºC for 1–2 hours. The reaction was stopped by inactivating the reverse transcriptase at 92ºC for 10 minutes. Newly synthesised cDNA was stored at -80ºC until ready to use in PCR reactions.

**O. PCR Primer design**

Primers were designed by hand towards the sequences of DIAP1 and cdc2 and prepared by Invitrogen Custom Primers (http://www.invitrogen.com). Primers were resuspended in sterile water to a stock concentration of 100µM, from which a 10µM working solution was prepared; both were stored at -20ºC. Primer design targeted areas of the DNA which were exclusive to the dsRNA sequence, so to observe only the priming of the target and not accidentally quantify the dsRNA which had been added to the cells. The list of specific primers used for each gene can be found in Appendix 1.

**P. Polymerase Chain Reaction**

PCR reactions were conducted using reagents provided in the Taq PCR Core Kit (QIAGEN). PCR reactions were prepared on ice to a total volume of 100µl as follows: 5µl 10X PCR buffer; 10µl 5X Q-Solution; 2–4µl $MgCl_2$ (25mM); 2µl dNTP mix (10mM each); 2µl each specific forward/reverse specific primer (5µl degenerate primers); 2µl template cDNA; 0.5µl Taq DNA polymerase (5U/µl); $H_20$ up to 100µl. PCRs were conducted using the Master-cycler Gradient Thermocycler (Eppendorf). Cycling conditions were kept constant for both PCRs. Initial DNA denaturation at 94°C (4 minutes), then 31 cycles of: Denaturation at 94°C (30 seconds), Annealing at 59°C (30 seconds), Extension at 72ºC (45 seconds) and the final extension at 72°C (5 minutes). Visualisation of the various PCR products was done via agarose gel electrophoresis. Gels were prepared by combining 1.5% agarose with 1X TBE (89mM Tris, 89mM boric acid, 2mM EDTA) and heated in a microwave to dissolve. For visualisation, 0.5µg/ml ethidium bromide was added to the liquid agarose before pouring into the gel cast. The DNA or RNA product was combined with MassRulerDNA Loading Dye (Fermentas) at a proportion of 1µl dye for every 5µl product and loaded into the wells of the agarose gel alongside the MassRulerDNA Ladder Mix (80–10,000bp fragments; Fermentas) to assess product length and to calculate approximate sample concentration. Gel pictures were taken using an Uvidoc gel documentation system and UviPhotoMW image analysis software.

**Q. Mitochondrial morphology assay**

Cultured S2 cells were resuspended in Serum-Free media and counted for dilution using a haemocytometer. Cells were to be plated at a concentration of 250,000 cells, which is a standard number for imaging. Six-well culture plates were prepared to contain an acid-treated coverslip upon which 2µg of 1µg/µl dsRNA working solution were pipetted. These dsRNAs were designed in-house by myself or J.L. Aspden and made by R.J. Phillips. In untreated cells, 2µl of milliQ H2O was pipetted instead of dsRNA. One millilitre of cells (250,000 cells/ml) were added to the coverslip and allowed to mix with the dsRNA by gentle swirling of the plate. After a 45 minute incubation at room

temperature, the cells + dsRNA were supplemented with Schneider's media containing 20% FBS and allowed to grow for 48 hours on the coverslip. After this time, MitoTracker-Red, diluted to 500nM in 10% FBS Schneider's media, was added to the wells for 1h. After removal of media, cells were fixed with 4% PFA, stained with DAPI for 10 minutes to label the nuclei and the coverslip was mounted on glass slides containing a drop of Vectashield.

**R. Bioinformatics Software**

Peptide Atlas (Desiere *et al.* 2006) Lists of peptide CDS coordinates with protein identifiers (FlyBase peptide ID) were downloaded from the Peptide Atlas database (http://www.peptideatlas.org). (Desiere *et al.* 2006) Peptide Sequences of smORFs were obtained from Ensembl (http://www.ensembl.org/biomart/martview). Prediction of transmembrane alpha-helices was performed using TMHMM (http://www.cbs.dtu.dk/services/TMHMM). Prediction of peptide-localisation motifs was conducted by TargetP (http://www.cbs.dtu.dk/services/TargetP/)

# Chapter III - Cellular characterisation of Hemotin

## Introduction

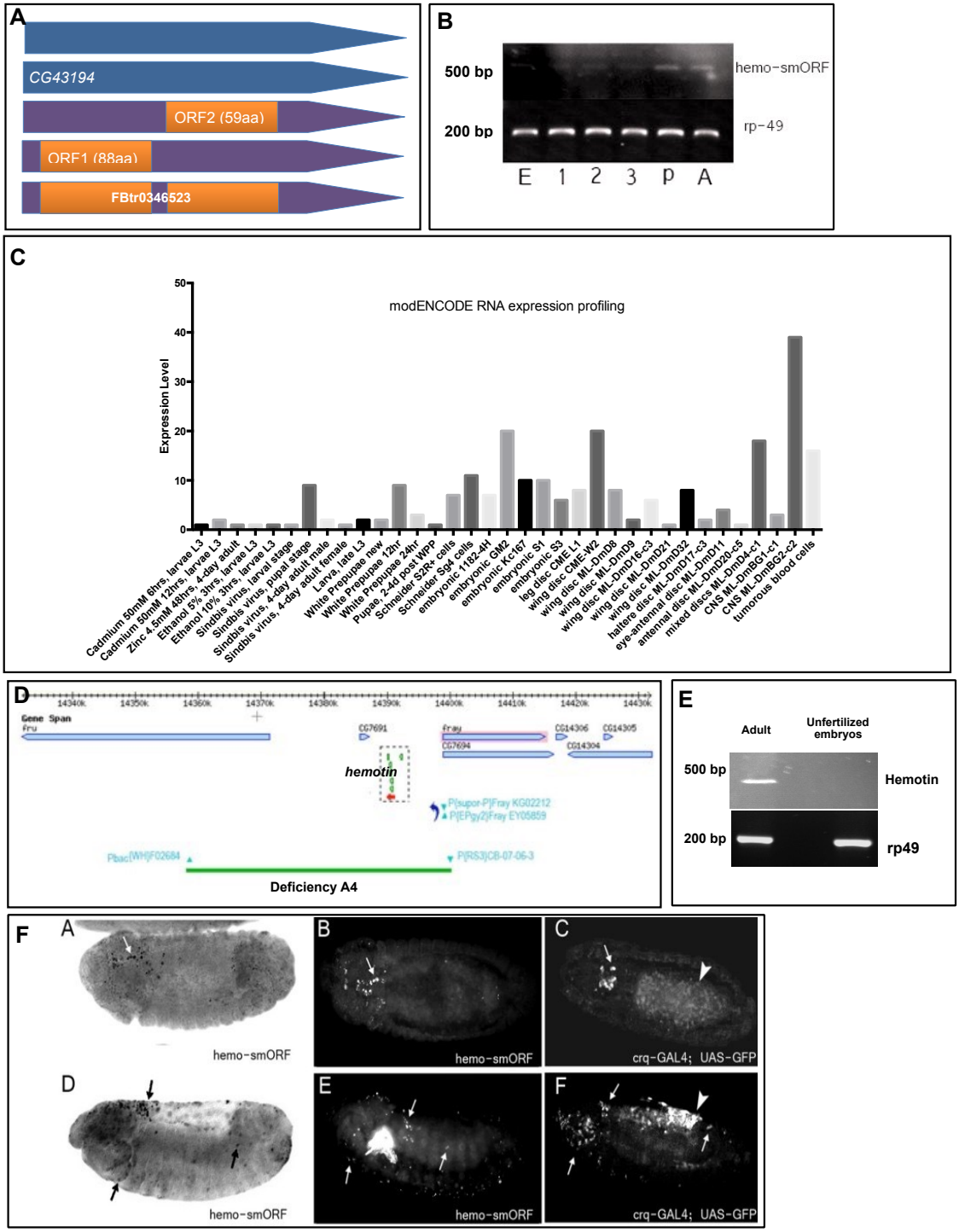
In proving the existence of small functional peptides, the detection of a peptide product encoded by a putative open reading frame is a crucial step in differentiating smORF transcripts from genuine non-coding RNA's. After the success of smORF gene characterisation with *tarsal-less* and *Sarcolamban*, a potential candidate was found on the *Drosophila* chromosome 3R. We named this gene *Swiss-smORF*, in homage to the detection of this putative peptide by a Swiss group in a genome-wide proteomics study of *Drosophila melanogaster*. The 'Swiss-prot' database was the result of one of the first genome-wide Mass Spectrometry studies conducted in the fruit fly (Brunner *et al.* 2007). With the promise of a detected protein, we began to map the various open-access data that is available regarding this stretch of DNA. *CG43194* and *CG43210* are now annotated as two individual gene models, but are actually a dicistronic smORF gene. There is evidence of a processed transcript for these genes, FBtr0346523, which contains two non-overlapping ORFs. The *Swiss-smORF* gene is approximately 1kb long, and encodes for two peptides of 88 and 59 amino acids in length (Figure 3.1A).

The 'model organism Encyclopaedia of DNA Elements' (modENCODE) project has compiled RNA-Sequencing data for several *Drosophila* experimental models, including various stages of animal development, tissue types, cell lines, and treatments. These experiments allowed for sequenced mRNA transcripts to be mapped to the genome in terms of their relative abundance in each of the conditions (Roy *et al.* 2010) (Li *et al.* 2014). Using the compiled data from various conditions, we are able to see that the transcript levels of *Swiss-smORF* are quite low overall (Figure 3.1B).

**N.B. Portions of this chapter have been published in Pueyo *et al.*, (2016) *PLoS Biology (in print)* and Sampson *et al.*, (2013) *Biology Open*.**

**Figure 3.1: Hemotin gene models and the expression of *Hemotin* in *Drosophila***

**A)** Hemotin is made of two smORFs that are transcribed as a dicistronic transcript encoding for 88aa (ORF1) and 59aa (ORF2) peptides. **B)** Semi-quantitative RT-PCR shows Hemotin is very lowly expressed at larval stages but is clearly detected at pupal and adult stages. Rp49 is included as a loading control. **C)** Summary of modENCODE RNA-Sequencing data showing expression of *hemotin* in various *Drosophila* developmental stages, cell lines and treatments. **D)** The genomic locus of *hemotin* and the deletion in the A4 deficiency *(Df A4)* line, which lacks the 5' ends of *fruitless* and *fray* and has a complete excision of *CG7691*. **E)** Semi-quantitative RT-PCR showing that there is no *hemotin* expression in unfertilised embryos compared to adult flies. Rp49 is used as a loading control. **F)** Pattern of expression of *hemotin* in stage 11 (A,B) and stage 15 (D,E) wild-type *Drosophila* embryos, as revealed by *in situ* hybridisation. *Hemotin* is expressed in in the head, amnioserosa, and dispersed along the body in both wide-field (A,D) and fluorescently-probed (B,E) images. *in situ* hybridisation of hemocyte-specific *Croquemort (Crq)* gene (C,F) showed similar distribution of Croquemort marker at stages 11 (C) and stage 15 (F) in the head, amnioserosa and along the body (arrows).


A
CG43194
ORF2 (59aa)
ORF1 (88aa)
FBtr0346523
B
500 bp
200 bp
hemo-smORF
rp-49
E 1 2 3 p A
C
modENCODE RNA expression profiling
Expression Level
50
40
30
20
10
0
Cadmium 50mM 6hrs, larvae L3
Cadmium 50mM 12hrs, larvae L3
Zinc 4.5mM 48hrs, 4-day adult
Ethanol 5% 3hrs, larvae L3
Ethanol 10% 3hrs, larvae L3
Sindbis virus, larval stage
Sindbis virus, pupal stage
Sindbis virus, 4-day adult male
Sindbis virus, 4-day adult female
Larva, late L3
White Prepupae new
White Prepupae 12hr
White Prepupae 24hr
Pupae, 2-4d post WPP
Schneider S2R+ cells
Schneider Sg4 cells
embryonic 1182-4H
embryonic GM2
embryonic Kc167
embryonic S1
embryonic S3
leg disc CME L1
wing disc CME-W2
wing disc ML-DmD8
wing disc ML-DmD9
wing disc ML-DmD16-c3
wing disc ML-DmD21
wing disc ML-DmD32
wing disc ML-DmD17-c3
wing disc ML-DmD11
wing disc ML-DmD20-c5
haltere disc ML-DmD4-c1
eye-antennal disc ML-DmBG1-c1
antennal disc ML-DmD20-c2
mixed discs ML-DmBG2-c2
CNS ML-DmBG2-c2
tumorous blood cells
D
Gene Span
Fru
CG7691
Fray
CG14306
CG14305
CG7694
CG14304
hemotin
P{supor-P}Fray KG02212
P{EPgy2}Fray EY05859
Pbac{WH}f02684
P{RS3}CB-07-06-3
Deficiency A4
E
Adult
Unfertilized embryos
500 bp
200 bp
Hemotin
rp49
F
A
hemo-smORF
B
hemo-smORF
C
crq-GAL4; UAS-GFP
D
hemo-smORF
E
hemo-smORF
F
crq-GAL4; UAS-GFP

The mRNA levels are normalised RPKM values (Reads per Kilobase/Million) on a scale of 1 to >1000. *Swiss-smORF* transcript is not detected in these experiments during embryonic stages. Very low levels of expression can be seen in conditions where the L3 larvae are treated with Cadmium, Zinc and Ethanol. *Swiss-smORF* levels increase with infection with the Sindbis virus, at the pupal stage (Figure 3.1C). Data from untreated flies shows that the expression of *Swiss-smORF* may be temporally regulated, with a peak of expression at the beginning of metamorphosis, 12h after puparium formation (WPP stage). This has been verified by RT-PCR, which showed faint expression of *Swiss−smORF* at L1, L2, L3 stages, followed by an increase of expression in the pupal and adult stages (Figure 3.1B). Comparably, there is a higher level of *Swiss-smORF* expression in *Drosophila* cell lines, such as those derived from the imaginal discs of the fly, which are the progenitor organs from which the adult structures (i.e. legs, wings, halters, eyes, antennae and gonads) are derived. All of these are locales where, over the course of *Drosophila* development, massive cell proliferation and cell death occur. Notably, expression is highest in the CNS derived cell line ML-DmBC2-G2, and in tumorous blood cells. All of this summarises the information that we knew before I began working on the project.

**Drosophila haematopoiesis during development**

*In situ* hybridisation of *Swiss-smORF* in *Drosophila* embryos showed a pattern of mRNA expression in the embryo where the signal was localised to the cephalic mesoderm in stage 11 (Figure 3.1F). From this result, we could observe that the areas of *Swiss−smORF* mRNA expression are similar to the expression of the *Croquemort (Crq)* receptor. This receptor recognises apoptotic cells for phagocytosis and is used as a marker for labelling haemocytes. *Crq* expression can be readily detected at late stage 11 of embryonic development (Franc *et al.* 1996) and *Crq-GAL4* is used as a haemocyte-specific driver in flies. *Swiss-smORF* expression is increased and more widespread in stage 15 embryos and again follows the expression pattern of GFP expressed using the *Crq-GAL4* driver. This led us to believe that *Swiss-smORF* may be expressed in the head region, and in haemocytes, which are the functional and cellular homologs of mammalian macrophages. At this time, the name of the gene encoding the *Swiss-smORF* peptide was changed to *Hemotin*, and that is how it will be depicted from now.

There are two haematopoietic ‘waves’ of blood cell differentiation during *Drosophila* development. The first occurs during embryogenesis and the other arises from the transient lymph gland at puparium formation (Holz *et al.* 2003) (Lanot *et al.* 2001). The main function of both embryonic and lymph gland haemocytes is to perform phagocytosis of apoptotic cells and their debris from developmentally programmed cell death, and to get rid of foreign pathogens (Tepass *et al.* 1994; Holz *et al.* 2003; Tzou 2002). Embryonic haemocytes are derived from the pro-cephalic mesoderm, in the presumptive head region, shortly after gastrulation (stage 10). Most cells in this compartment go through several rounds of mitosis to give rise to recognisable pro-haemocytes (Tepass *et al.* 1994). During germ-band retraction (stage 12), haemocytes begin to migrate from the cephalic region to populate the entire embryo (Tepass *et al.* 1994; Cho *et al.* 2002; Evans *et al.* 2010). Haemocytes can first be distinguished from pro-haemocytes in the head during Central Nervous System (CNS) formation (early stage 13), where they are required to clear up the cellular debris around the developing nervous system. In haemocyte depleted embryos, the CNS fails to condensate along the anteroposterior axis (Defaye *et al.* 2009). Haemocytes in these areas start to show inclusions, or vacuoles, containing the leftovers of apoptotic cells, indicating that the haemocytes have become phagocytic. Almost all (~90%) embryonic pro-haemocytes change into macrophages by stage 15 of embryogenesis (Tepass *et al.* 1994). A remaining ~5% of pro-haemocytes found near the embryonic head, develop into crystal cells (Lanot *et al.* 2001).

*Vascular endothelial growth factor (VEGF)* genes are expressed along the developmental migration paths of the haemocytes, and the VEGF receptor (VEGFR) is expressed in *Drosophila* haemocytes. The three isoforms of the VEGFR play a redundant, yet essential role in the migration of haemocytes (Cho *et al.* 2002). The ability of haemocytes to migrate at embryonic stages is not dependant on signalling from apoptosis but rather, developmental migration of haemocytes is mediated by the expression of VEGF, which happens to coincide with areas of embryonic apoptosis (Cho *et al.* 2002). In *Drosophila,* haemocytes are required for normal embryonic development as inducing apoptosis in haemocytes using a haemocyte-specific GAL4 resulted in 94% lethality of embryos before the first Larval instar stage (Defaye *et al.*

2009).

Before hatching into larvae, embryos develop a presumptive lymph gland along the dorsal vessel, close to the cells that will constitute the heart of the fly. The lymph gland is derived from a mesodermal origin during embryogenesis. The lymph glands have two pairs of lobes, and the first to appear are the anterior/primary lobes. At early L3 stage, the posterior or secondary lobes begin to develop (Crozatier and Meister 2007). The second haematopoietic wave occurs at the end of larval stages when the pro-haemocyte group of cells found in the posterior lobe of the lymph gland begin to differentiate into macrophages. The mature macrophages are released upon pupation, which is followed by the disintegration of the lymph gland (Jung *et al.* 2005). The schematics in Figure 3.2A depict the process of haematopoiesis in *Drosophila.* The lymph gland is a source of haemocytes primed for the phagocytosis of apoptotic material released during metamorphosis, and the later requirement for the adult fly immunity (Lanot *et al.* 2001). However, lineage tracing reveals that both embryonic and lymph gland haemocytes are found in the adult haemolymph (Holz *et al.* 2003). Our initial observation of *Hemotin* mRNA in the presumptive head region during embryogenesis and the increased expression of *Hemotin* during L3 and pupal stages, fits into the timing of important haemocyte-related events such as differentiation.

**The haematopoietic role of ecdysone**

In the early 1900's, it was discovered that insect moulting, the mechanism by which insects develop and grow, is a hormone-dependent process. This hormone was discovered to be a pentahydroxysteroid (alpha-ecdysone). The conversion of the pentahydroxysteroid to a hexahydroxy form *in vivo*, resulted in the later identification of Beta-ecdysone (Garen *et al.* 1977). Ecdysones have been discovered in many other insects that use moulting as a mechanism for growth. Ecdysone is involved in the progression of *Drosophila* developmental stages (Yamanaka *et al.* 2013), and is required for proper embryonic development (Kozlova and Thummel 2003) and the larval to pupal stage transition (Figure 3.3B) (Thummel 1996). Waves of ecdysone expression coincide with polytene chromosomal puffing, which is a sign of gene transcriptional activation. These genes were found to encode the Broad-Complex (BR-C) transcription factors E74 and E75, along with the Ecdysone receptor gene (EcR).

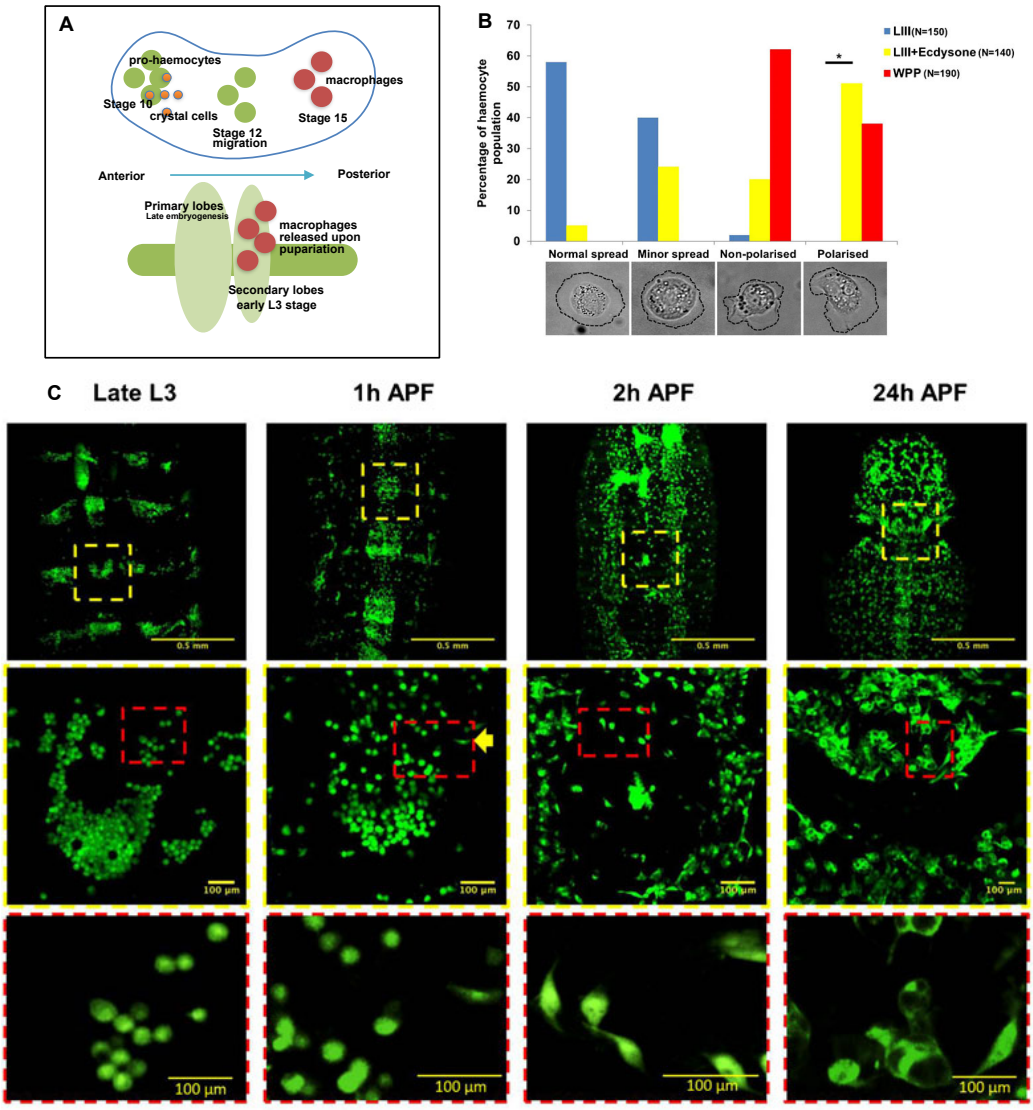



**Figure 3.2: Developmental stages of haematopoiesis and ecdysone activation of haemocytes** **A)** Schematic of the differentiation of haemocytes in the *Drosophila* embryo (top) and larval lymph gland (bottom). The haemocytes migrate as they differentiate from the anterior to the posterior of both embryo and lymph gland **B)** *Ex vivo* studies showed a distinct change in cell morphology during the transition from late L3 to white pre-pupal stages from 'normal' spread at L3 (blue bars) to 'polarised' when exposed to endogenous ecdysone at WPP stage (red bars). This activation can be replicated by Ecdysone treatment of L3 haemocytes (yellow bars) *(Sampson CJ, Amin UA, Couso JP, 2013, Biology Open)*. **C)** I conducted *in vivo* imaging of GFP-expressing *Drosophila* haemocytes during L3, 1 hour, 2 hours (White Pre-Pupae, WPP) and 24 hours After Pupal Formation (APF). Haemocytes show a change in morphology from rounded (L3, 1hAPF) to activated, polarised shape (2hAPF) as Ecdysone levels increase at the WPP stage *in vivo*. 24hAPF haemoctes are full of phagocytosed material.

The three isoforms of the Ecdysone Receptor (EcRA, EcRB and EcRC) can function as receptors for hormonal signalling after they undergo hetero-dimerisation with the Drosophila RXR homolog Ultraspiracle (Usp) (Thummel 1996). Tissue-specific expression of ecdysone-responsive transcription factors and EcR isoforms occurs to regulate the programmed cell death or autophagy pathways in various larval tissues that are not needed in the adult fly. These include the salivary glands and segmental muscles, which are specific to the feeding and movement requirements of the larvae (Jiang *et al.* 1997). Targeted elimination of specific cells in the ommatidia allows the correct differentiation of the extremely complex adult Drosophila eye, and in a similar manner, specific CNS cells expressing EcRA are removed at the end of the larval stages (Truman *et al.* 1994). As mentioned above, haemocytes are responsible for the removal of cellular debris arising from the histolysis of the larval organs.

It has been shown that haemocytes extracted from mid-third instar, wandering stage and late-third instar larvae (L3) are rounded in shape, whilst those extracted from early white-pre pupae (WPP) have a flattened appearance when they adhered to a substrate (Lanot *et al.* 2001). The L3 to WPP transition is coincidental with a large peak in Ecdysone (Bate 2009). To test whether the change that occurs during this phase is due to the action of this hormone, ecdysone was injected into mid-third instar larvae and after five hours; the larval haemocytes produced the same morphology as WPP stage haemocytes (Lanot *et al.* 2001). We have further discerned that when L3 haemocytes are cultured with ecdysone *ex vivo,* they do begin to show the characteristics of WPP haemocytes (Sampson *et al.* 2013). *Ex vivo* cultured wild-type L3 stage haemocytes incubated with ecdysone for three hours changed the characteristic morphologies of L3 haemocytes from sessile to a motile state; with comparable migration velocities and cytoskeletal re-arrangements to those of haemocytes extracted from wild-type WPP flies (Figure 3.2B) EcR dominant-negative L3 haemocytes do not show the this change when cultured with ecdysone *ex vivo,* as seen in wild-type cells (Sampson *et al.* 2013). The behaviour of haemocytes in *ex vivo* culture recapitulates what can be observed *in vivo* (Figure 3.2C). I have previously shown that haemocytes found attached to the inside of the larval integument at the end of L3 are rounded in shape, and show little movement out of the patches and into circulation. One to two hours after puparium formation

(APF), the haemocytes that were previously attached to the integument began to show characteristics of motility such as protrusion formation, followed by dispersion from the dorsal patches, as observed through fluorescent live-imaging of WPP haemocytes. I also showed that haemocytes expressing dominant-negative forms of EcR do not migrate out of the dorsal patches *in vivo*. This led us to conclude that larval haemocytes are activated into motility upon puparium formation by Ecdysone and this action is mediated through the Ecdysone receptors. Twenty-four hours APF, cells found in the dorsal region of the thorax and abdomen show large blebs in their cytoplasm. We term these 'vacuoles' as they resemble large intercellular compartments that are full of material, similar to those found in plant cells. The specific temporal expression of Hemotin can be correlated to the highest peaks of ecdysone hormone expression level, known to trigger apoptosis at L3 to pupal transition, creating requirement for phagocytosis by haemocytes in late larval and early pupal stages.

**The role of haemocytes during *Drosophila* development**

If Hemotin performs a functional role in haemocytes, it becomes necessary to understand the role of haemocytes in *Drosophila* immunity. In the wild, *Drosophila* spend most of their lives inside and around rotting fruit. Eggs are embedded inside, providing hatched larvae with an instant food source that remains a reliable, nutrient-rich habitat for adult flies. It is no wonder then that the fruit fly has evolved a complex immune system to prevent infection by the various microorganisms that may be found there. Like most insects, *Drosophila* have a difficult-to-penetrate epidermal cuticle, and like most animals, a lower pH in the gut that is inhospitable for microorganism proliferation. *Drosophila* have an open circulatory system and its rudimentary 'heart', the dorsal vessel, pumps haemolymph around the body. The haemolymph contains the cellular arm of the Drosophila immune system, in the form of 3 main cell types: lamellocytes, crystal cells and plasmatocytes/haemocytes. Each of these cell types perform specific roles in the follow-up to infection (Tzou *et al.* 2002).

Wounding the fly triggers a melanisation response by the crystal cells, followed by lamellocyte or haemocyte recruitment and the subsequent production of anti-microbial peptides (Hoffmann *et al.* 1999; Ulvila *et al.* 2011a). Lamellocytes are highly

specialised cells that are used in the encapsulation of parasitic wasp eggs or very large molecules (Tzou *et al.* 2002). Therefore, for the purpose of clarity, we maintained focus specifically on haemocytes, which are the predominant form of macrophages/ phagocytes of the *Drosophila* immune system (Lemaitre and Hoffman 2007). *Drosophila* haemocytes are highly motile phagocytic cells that are present at all stages of the fly life cycle and represent ~90% of the cellular component of the insect's immune system during post-embryonic life stages (Lanot *et al.* 2001).

In mammals, blood cells are derived from stem cells through expression of varying combinations of regulators and specific transcription factors. Similarly in *Drosophila*, the mammalian GATA transcription factor homolog Serpent (Srp) is required for the development of embryonic pro-haemocytes as well as the embryonic gut. Mature haemocytes could not be found in embryos lacking the *srp* gene, which also showed impaired pro-haemocyte differentiation (Rehorn *et al.* 1996). Another key player in mammalian blood cell differentiation is the protein AML1, which is also required for haematopoiesis. Lozenge (lz), the *Drosophila* AML1 homologue, is required for the differentiation of crystal cells, and is expressed in a small subset of Srp expressing pro-haemocytes during stage 11 of embryogenesis, when haemocyte differentiation begins. Rather than migrating around the embryo, crystal cells cluster in the head. Lz expression is also maintained in a subset of cells in the anterior lobes of the larval lymph gland. Srp is required for, and is upstream of Lz expression during haematopoeisis. Pro-haemocyte differentiation into macrophages requires the transcription factor *glial-cells-missing* (Gcm) (Lebestky *et al.* 2000). The mammalian Friend-of-GATA (FOG) protein homolog, U-shaped (Ush), is responsible for negative regulation of crystal cell production in *Drosophila* (Fossett *et al.* 2001) and expressed in macrophages.

The production of anti-microbial peptides (AMP's) from the fat body is another mechanism of the innate immune response that is conserved from *Drosophila* to mammals, with up to 400 peptides identified in insects, plants and humans (Hoffmann 1999). AMPs are released as a result of a wound or infection. Oral infection of flies during the larval stage does not elicit an AMP response in the absence of functional haemocytes, though this observation is not repeated in infected adult flies (Charroux

and Royet 2009; Defaye *et al.* 2009). The various parallels between *Drosophila* and vertebrate immunity make it more interesting to investigate whether there is a functional role of this SEP in immunity.

**Phagocytosis and intracellular trafficking**

The uptake of material from the cellular environment through phagocytosis is a highly conserved mechanism from *Drosophila* to mammals (Hoffmann *et al.* 1999). The recognition of foreign particles by the haemocyte is the first step of many signalling and cytoskeletal re-modelling events that take place in this process (Ulvila *et al.* 2011a). Phagocytosis mediates uptake of particles that are ≥0.5μm in size. One of the receptors involved in the recognition of foreign particles is the CD36 scavenger receptor homolog Crq, which is expressed in all embryonic haemocytes (Franc *et al.* 1999). This is only one of the pattern recognition receptors (PRRs) involved in the recognition of microbial-associated molecular patterns (MAMPs) which are unique to various micro-organisms. Binding of these receptors to their respective ligands initiates downstream signalling; which induces expression of AMPs in larvae, Actin-associated proteins, endocytic machinery and regulators of intracellular vesicle trafficking (Ulvila *et al.* 2011a). In a similar manner to pathogens, apoptotic cells release a phylogenetically conserved 'eat me' signal in the form of a Phosphatidylserine (PS) which is externalised by the dying cell (VanDenEijnde *et al.* 1998; Fullard *et al.* 2009). In *Drosophila,* recognition of apoptotic cells is conducted by Draper and similar EGF-repeat containing proteins Nimrod and Eater; which are also necessary for the recognition of various types of bacteria (Fullard *et al.* 2009). After particle recognition, internalisation of particles is mediated by the probing movement of the plasma membrane called membrane ruffling, which can be observed in both embryonic and post-embryonic haemocytes. These movements facilitate the binding of receptors to their target. Membrane ruffling is mediated by the GTP hydrolase enzymes (GTPases) of the Rho family (Cdc42, Rac1 and Rac2). These GTPases activate the WASP and SCAR/WAVE family of proteins, which in turn stimulate Arp2/3. Arp2/3 nucleates actin and promotes filament formation to produce the phagocytic cup (Shim *et al.* 2010). These are the main proteins known to be responsible for the formation of membrane protrusions, but there is not that much understood about the membrane fusion events that actually mediate the internalisation

of particles (Shim *et al.* 2010; Botelho and Grinstein 2011). The internalised membrane-bound particles, now called phagosomes, undergo a series of maturation events that change the membrane composition as well as the contents. As the phagosome matures from an early-endosome to a late-endosome and then a phago-lysosome, the organelle becomes more acidic with enzymes, thus leading to the degradation of the phagocytised particles. These organelles display different Rab GTPases on the membrane surface which act as recruitment factors for various effector proteins which affect the internal pH and movement of the organelle through the cell (Stein *et al.* 2003).

**Avenues of investigation into the function of Hemotin**

There is significant homology between mammalian and *Drosophila* innate immunity, and both insect and vertebrate macrophages function in a similar fashion. Phagocytosis is an important cellular function performed by these cells in all organisms in which they are found. Our preliminary data on Hemotin required the investigation of several potential mechanisms that this smORF peptide could be involved in. We hypothesised that Hemotin is expressed in haemocytes during the first wave of haematopoiesis in embryos, based on the patterns of mRNA expression in embryos, though we did not know the endogenous peptide's localisation. The peak of *hemotin* RNA in *wild-type* and Sindbis virus infected pupae, coincides with a the second haematopoietic wave in the fly. As Hemotin is a small peptide, it could be involved in immunity as an AMP or perhaps act downstream of ecdysone signalling to activate haemocytes into motility and phagocytosis. It could even be involved in the mechanism of phagocytosis. The possibility that smORF peptides may be involved in these essential cellular functions is exciting, as no smORFs have been previously associated with these functions.

*Hemotin* is not detected by RT-PCR in unfertilised embryos (Figure 3.1E) indicating that it is not a maternally deposited. Therefore, we went on to generate a deficiency line over the genomic region of the *hemotin* gene using two P-element reversions that are found between the 5' ends of the *fray* and *fruitless* alleles in the genome. This allowed the excision of the entire *hemotin* gene, as well as *CG7691*, which is annotated to contain a putative Zinc finger domain with predicted functions in metal ion and nucleic acid binding. The genomic excision also removes the 5' ends of the *fray* and *fruitless*

genes (Figure 3.1D). The *hemotin* null mutant flies are viable and *wild-type* looking morphologically, but male sterility exists in the deficiency line, possibly due to the excision of the 5' end of the *fruitless* gene (Demir 2005). We depicted this null mutant for *hemotin* as *Df A4* to be used in a reverse genetics approach in the characterisation of *hemotin*. The aim of the experiments described in this chapter were to use various fly lines, including the *hemotin* null mutant *Df A4,* to elucidate the function of Hemotin SEP in a single-smORF study. This reverse genetics approach has been previously used by our laboratory to characterise smORF protein function in flies after removal of the *tarsal-less* and *sarcolamban* genes.

# Results

**Detecting Hemotin peptide in embryos and post-embryonic haemocytes**

One of the first steps in characterising a putative ORF is to obtain proof of translation. The standard technique of antibody staining (immunohistochemistry, IHC) is used to detect the translated peptide. Preferential binding of a specific antibody raised against the antigen, serves as proof of protein translation of the bait protein. Testing the antibody in *wild-type* flies would technically show where the endogenous peptide is being made. A polyclonal antibody was generated against the Hemotin peptide (peptide sequence: RISITPGCSMVATTKLTHSRNSVDIY). There were two versions of the antibody, a Serum Antibody (SAB), which is the blood serum extracted from the animal 28-days after the injection of the peptide from the animal (rabbit). Some of the serum is subjected to Affinity/Immuno-purification (IP) and checked by ELISA, resulting in a concentrated antibody which is specific to the C-terminus sequence of Hemotin (Eurogentec).

Labelling the endogenous expression of Hemotin peptide in fertilised embryos would also give a better idea that mRNA expression (as observed previously by *in situ* hybridisation) as to which tissue or cell type where Hemotin is functioning. We had previously postulated that Hemotin is expressed in haemocytes from evidence of expression at the mRNA level, but we required confirmation that the protein is also found in these cells to conduct a functional study, not being exported or transported to some other tissue. I used a fly stock carrying SrpGAL4 driving the expression of GFP, *(w; SrpGAL4>GFP; Df A4/TM3-YFP)* to label embryonic haemocytes. Embryos were collected 8-16 hours after egg laying, which encompasses stages of development 12-16, at which haemocytes have differentiated and migrated around the embryo. The embryos contain three genotypes on chromosome III *Df A4/TM3-YFP, Df A4/Df A4* and *TM3-YFP/TM3-YFP.* All the haemocytes in this fly line express GFP, due to the homozygous *SrpGAL4>GFP* construct on chromosome II. Embryos with the *TM3-YFP* balancer can be seen by the expression of yellow fluorescent protein in the embryos. Dfd-GMR drives expression of YFP in the presumptive eye and proboscis of the embryo(Le 2006). *Df A4/TM3-YFP, and TM3-YFP/TM3-YFP* embryos were used as the *wild-type* sample,

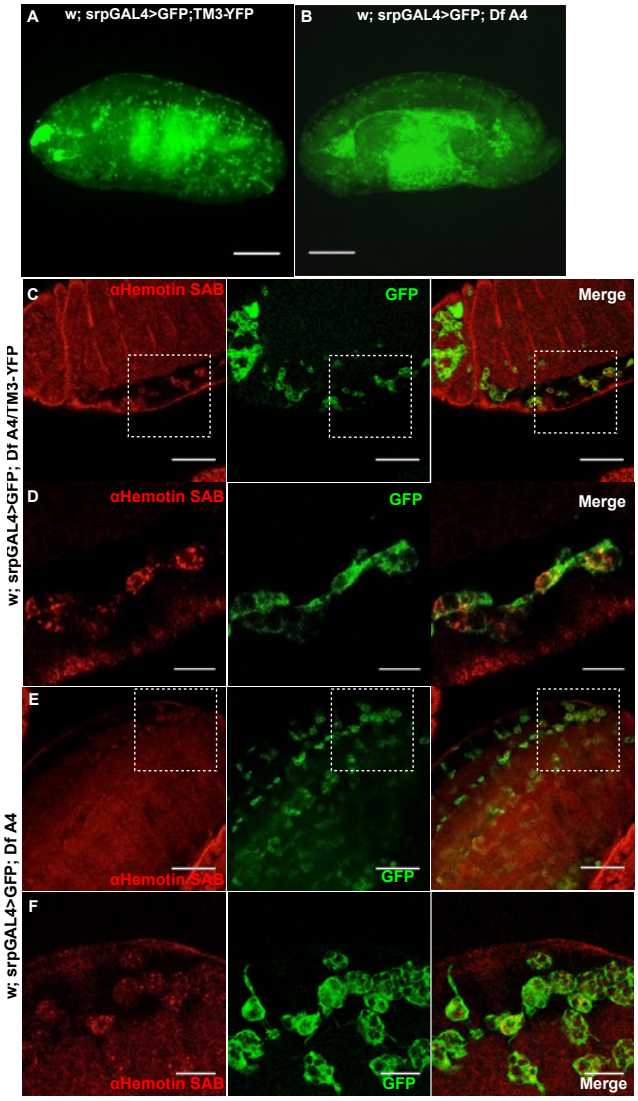



**Figure 3.3: Testing the anti-Hemotin Serum Antibody in embryos**

Embryos laid from flies with genotype w; *srp*GAL4>UAS-GFP; *DfA4*/TM3-YFP labeled with anti-GFP (FITC secondary) as seen in **A)** wild-type *DfA4*/TM3-YFP or TM3-YFP/TM3-YFP embryos and **B)** *hemotin* deficient *DfA4/DfA4* embryos. These can be differentiated by the difference in YFP expression in the presumptive eye and proboscis and the *hemotin* deficient embryos serve as an internal negative control to test the antibody Scale bar = 100µm. **C)** anti-Hemotin Serum-antibody (SAB) (Rhodamine secondary) localises in a punctate fashion, found within GFP-labelled hemocytes in *wild-type* embryos. Scale bar = 50µm **D)** 5x zoom of cells from panel C. Scale bar = 5µm. **E)** anti-Hemotin Rhodamine signal can also be seen within haemocytes in *hemotin* deficient *(Df A4)* mutant embryos **F)** 5x zoom of haemocytes from panel E.

and would contain the coding sequence for *hemotin*. The homozygous null (*Df A4)* embryos served as an internal negative control in these samples. The difference between *wild-type* and negative control embryos is shown in Figure 3.3A,B. Using a combined pool of *wild-type* and negative control embryos also ensures that the immunohistochemistry protocol is consistent between the two conditions. Using the SAB it could be seen that there is strong labelling of puncta within *wild-type* haemocytes, as compared to signal from the mesoderm and non-GFP expressing areas (Figure 3.3C,D). However, in *Df A4* embryos, there is also punctate signal within haemocytes. The signal in the null-mutant embryos is of comparable strength and quality as the *wild-type* condition (Figure 3.3E,F). Firstly, this result made us question the specificity of the antibody generated against Hemotin but it was also considered that labelling of Hemotin in embryos may have been difficult as there may not be high enough levels of expression at embryonic stages (modENCODE data). The main peak of Hemotin expression in *wild-type* flies is seen at L3 to pupal stages, but this requires accessing haemocytes at these stages for staining.

As the SAB showed signal in the Df A4 embryos, we decided to use the Immunopurified (IP) version of the antibody for subsequent experiments as it may be more specific due to higher purity. When the IP antibody was tested in embryos, it showed some signal in *wild-type* embryonic haemocytes (Figure 3.5A,B). As described with the SAB, a lot of fluorescent background can be seen in the mesoderm, and in areas where haemocyte-specific GFP is not being expressed. Despite trying different staining protocols, signal level is comparable in both GFP-labeled and non-labeled regions. This made it hard to tell whether the IP was any better than the SAB. L3 and WPP haemocytes were bled from animals to conduct immunohistochemistry directly on the haemocytes to reduce background, and in hope that there would be more Hemotin expressed at these life stages. Results from IP staining of *ex vivo* haemocytes revealed Punctate red signal in the *wild-type, Or-R* L3 haemocyte cytoplasm (Figure 3.4A). This pattern of signal is reproduced in *Or-R* and TM3-YFP WPP haemocytes using the IP antibody (Figure 3.4B,C) Unfortunately, in haemocytes bled from *Df A4* L3 and WPP animals (Figure 3.4D,E), the same pattern of signal can be observed in the red channel.

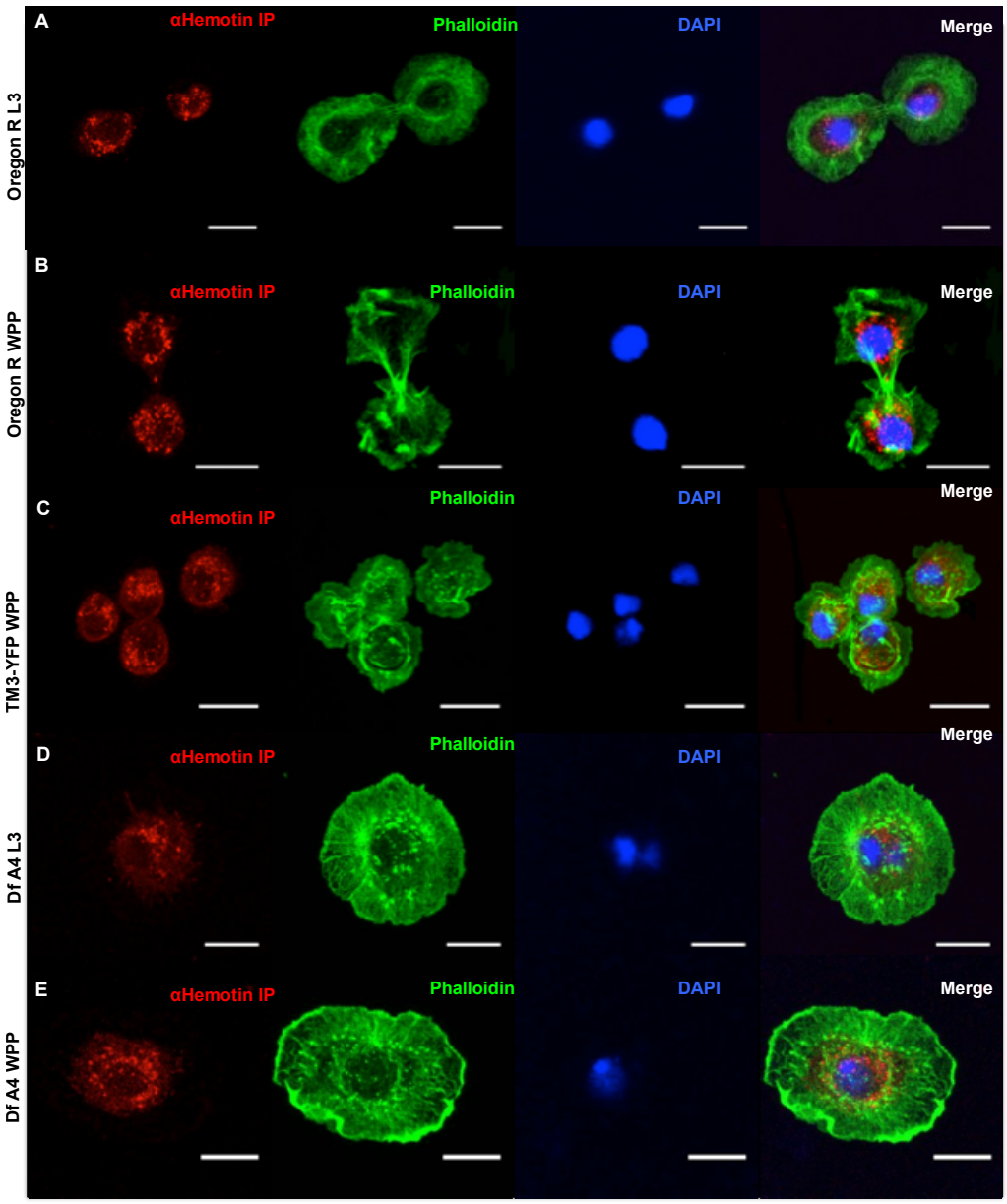



**Figure 3.4: *Ex vivo* staining of larval and pre-pupal haemocytes with anti-Hemotin IP antibody** **A)** *Wild-type (Oregon-R)* L3 larval haemocytes stained with immune-purified (IP) anti-Hemotin antibody (labelled with Rhodamine). The cellular F-actin is labelled with Phalloidin-FITC and nuclei are labelled with DAPI (Blue) **B)** *Oregon R* white pre-pupal (WPP) haemocytes with same treatment as panel A **C)** *Df A4/TM3-YFP* WPP haemocytes. Anti-Hemotin IP labelling shows a punctate signal in the cytoplasm in wild type cells. *Df A4* homozygous haemocytes stained with anti-Hemotin IP antibody is similar to wild type in *Hemotin* deficient hemocytes at **D)** L3 and **E)** WPP stages (Scale Bars = 5μm)

In the case of the characterisation of another smORF, detecting the peptide with a specific antibody led to similar issues as described above. It was found that Tal could only be detected by anti-Tal antibody when expressed as a fusion protein with a Glutathione S-transferase (GST) tag (J.I. Pueyo, personal communication), so we decided to see whether the same would occur in the case of Hemotin tagged with a larger epitope. Using a GFP tagged fusion construct of Hemotin-ORF1, expressed with a hemocyte-specific GAL4 (Figure 3.5E), I retested the IP antibody in the hope it will detect the larger, GFP-fused Hemotin-ORF1 peptide, hereafter referred to as Hemotin-GFP. *UAS-Hemotin-GFP* is a fusion construct in which the GFP sequence (devoid of a start codon) has been cloned into the hemotin-Full-Length construct (containing both ORFs), downstream and in-frame with *hemotin* ORF1, which lacks a stop codon (Pueyo *et al.* 2016). In both L3 and WPP haemocytes expressing *UAS-Hemotin-GFP,* there is very little overlap by the signal derived from anti-Hemotin antibody to Hemotin-GFP, in fact, they seem to be mutually exclusive of each other (Figure 3.5C,D). This result, combined with the fact that there is signal from the IP in *Hemotin* null mutant haemocytes, led me to discontinue use of the Hemotin antibody as it was nonspecific.

**Colocalisation of Hemotin-GFP to organelle markers**

Although I was not able to determine endogenous Hemotin localisation in the embryo and specifically in haemocytes, the GAL4-UAS system allowed the targeted expression of UAS-Hemotin in specific tissues, including haemocytes, with well-characterised haemocyte GAL4s. The detection of GFP-tagged construct of Hemotin provides proof of translation of Hemotin, because the GFP is only translated if Hemotin is being translated due to the fusion of the GFP sequence. The expression of detectable GFP would help to determine the subcellular distribution of the fusion peptide by co-labelling various organelles or cytoskeleton in haemocytes. This would provide information about where Hemotin may be functioning. It has been described that most of the organelles share a significant part of their inner membranes. The juxtaposition of Endoplasmic reticulum (ER) membranes with mitochondrial outer-membranes are called mitochondria-associated membranes (MAM) (Vance *et al.* 1997). It is also common knowledge that peroxisomes and other transport vehicles bud off from the ER

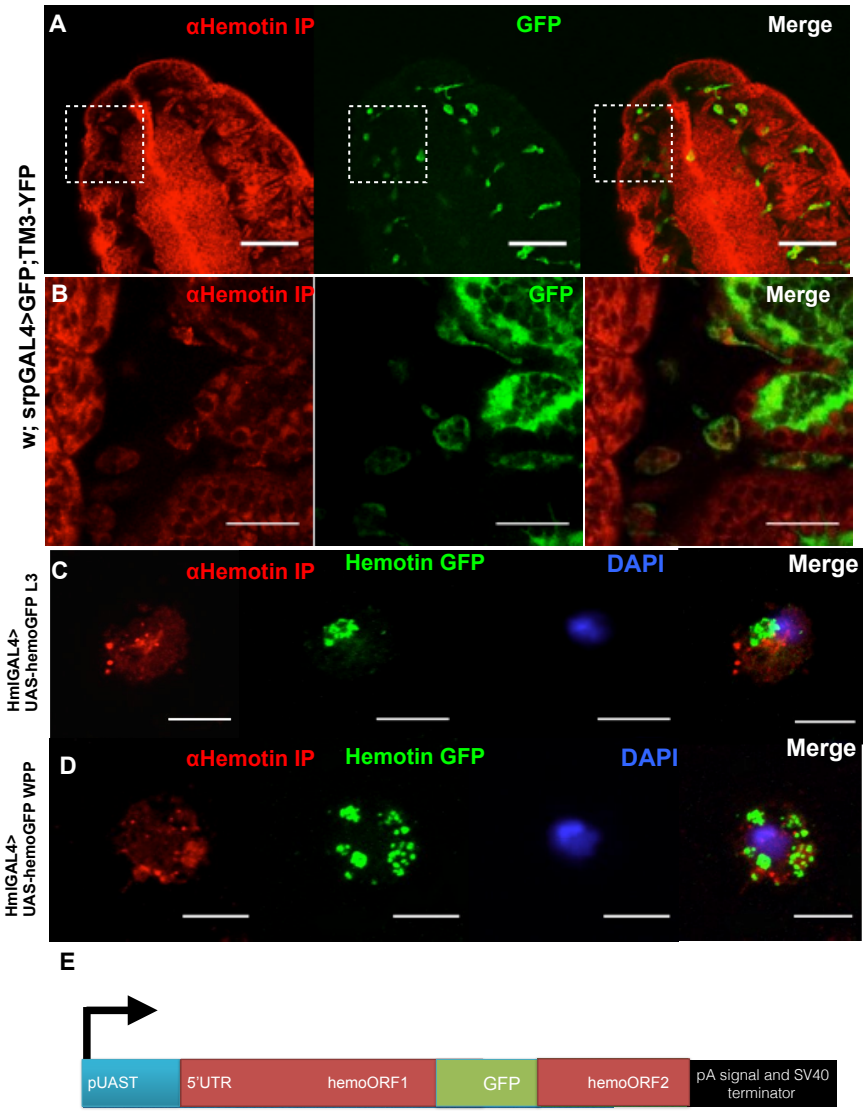



**Figure 3.5: anti-Hemotin IP antibody staining in embryos and *ex vivo* Hemotin-GFP haemocytes A)** Staining of Embryos with IP Hemotin Antibody (Rhodamine) shows extensive background staining of the mesoderm that does not overlap with *srpGAL4>UAS-GFP* marked hemocytes. Scale Bar=50µm. **B)** 5x zoom of panel A. Scale bar = 5µm. **C)** *UAS-Hemotin-GFP* expression driven by haemocyte-specific *Hml-GAL4* in **(C)** L3 and **(D)** white pre-pupal haemocytes stained with anti-Hemotin Antibody (Rhodamine) and DAPI (Blue). Rhodamine and GFP signal show very little overlap. Scale bars = 5µm. **E)** Schematic of the UAS-Hemotin-GFP construct expressed in flies using Hml-GAL4

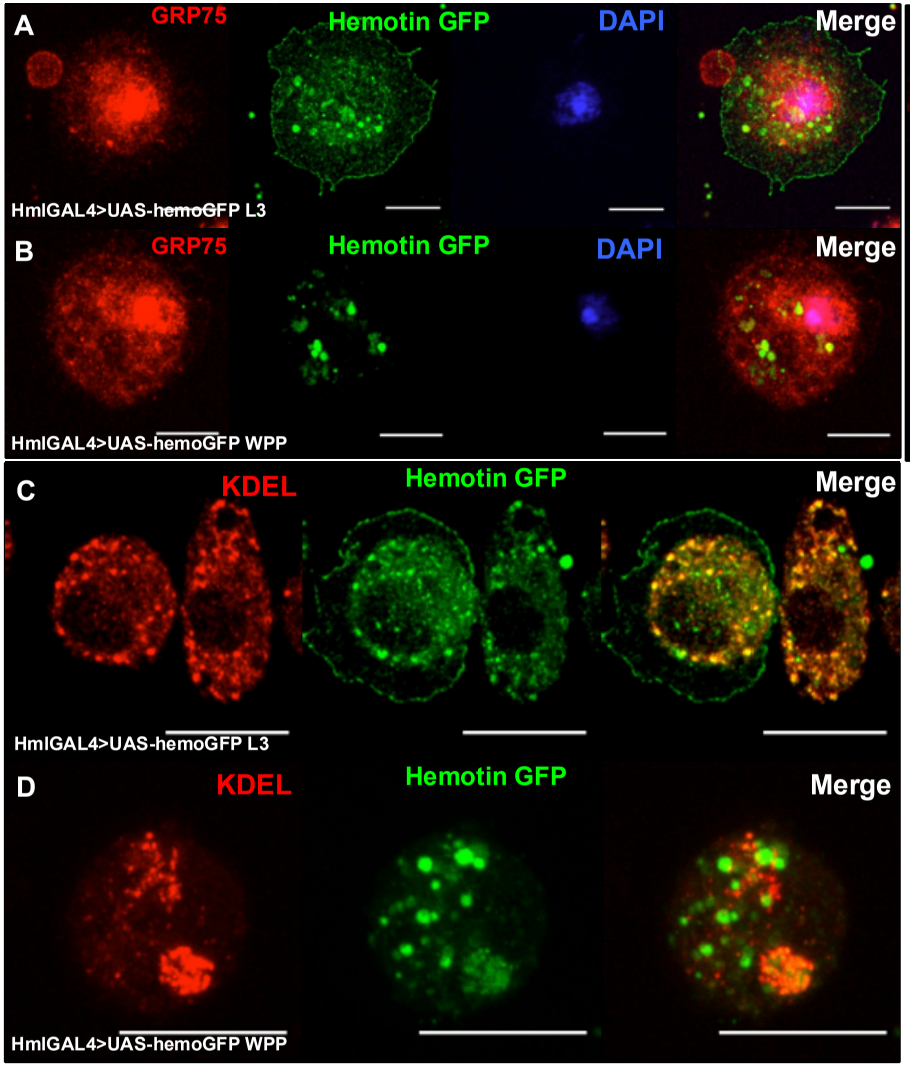
and Golgi (Lodish 2008). Since the expression of GFP was in a network and punctae all over the cytoplasm, we decided to look at organelle markers of the inner membrane system: the mitochondria/anti-GRP75 (Santa Cruz), anti-KDEL (DSHB) and Lysotracker-RED (Molecular Probes). The study was conducted on *ex vivo* haemocytes from L3 and WPP flies, which enabled us to combine the power of fly genetics with high high resolution microscopy in the hope of elucidating the cellular function of Hemotin peptide.

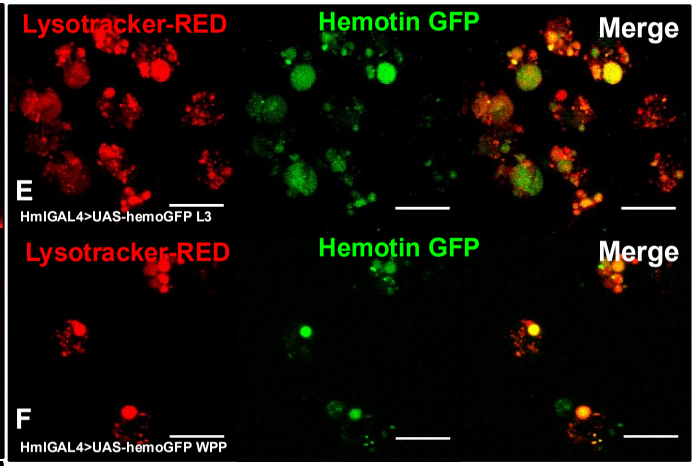
Figure 3.6A,B shows staining in L3 and WPP haemocytes with the mitochondrial antibody GRP75. There could not be any significant overlap seen between the mitochondria and Hemotin-GFP. We then tried to label KDEL which is a protein motif that labels the membranes of ER, Golgi and transport vesicles between the two organelles, as well as vesicles targeted for the ER from the plasma membrane (Stornaiuolo *et al.* 2003). This staining revealed a high level of overlap between Hemotin-GFP and KDEL in L3 haemocytes (Figure 3.6C,D). The amount of overlap between Hemotin-GFP and KDEL in WPP haemocytes is lower than what is seen at L3, but this result allows us to conclude that Hemotin is found on the ER, Golgi and intracellular transport organelles, such as those that bind to lysosomes and phagocytised material, thus making it a viable candidate for having a function in intracellular trafficking. In order to visualise Hemotin-GFP signal in relation to Lysotracker, which labels endo-lysosomal vesicles with a low pH, I had to live-image haemocytes from L3 and WPP animals. The reason for this was because Lysotracker permeates the membranes of live cells, and aldehyde fixation is not feasible subsequent to Lysotracker staining. This lack of fixation makes it difficult to amplify fluorescent signals, which in turn leads to an overall decrease in the strength of fluorescent signal that can be captured, as the microscopy parameters have to be kept at a lesser intensity to avoid bleaching the live sample. Despite this limitation, I was able to capture images of Hemotin-GFP expressing haemocytes, and the GFP signal in relation to Lysotracker-Red, which led us to conclude that Hemotin does indeed colocalise quite strongly to Lysosomal compartments (Figure 3.6E,F).

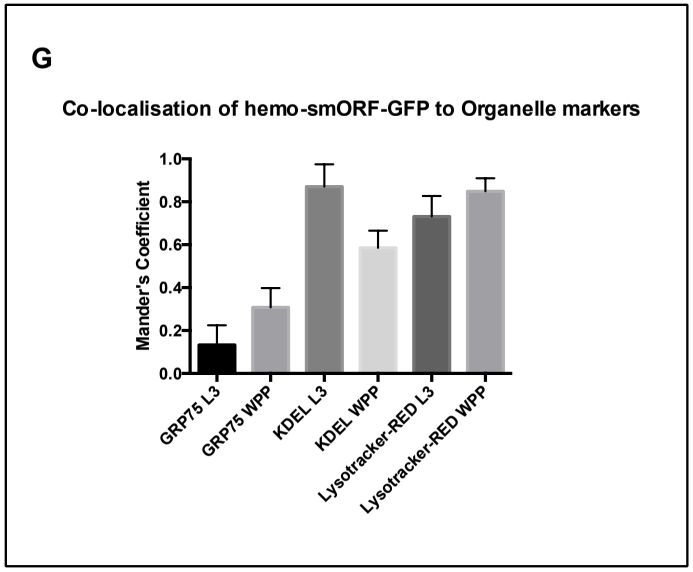
In order to make the result more convincing, I decided to quantify signal overlap between the organelle markers and Hemotin-GFP. Colocalisation between pixels can be

**Figure 3.6: Colocalisation of Hemotin-GFP with Organelle markers**

Hemocytes bled from *HmlGAL4>UAS-Hemotin-GFP* L3 and WPP animals are co-stained with **(A, B)** Mitochondrial marker Grp75 **(C, D)**, ER marker KDEL and **(E, F)** live-imaged using Lysotracker-RED dye to label lysosomes. **G)** Quantification of Hemotin-GFP signal with organelle markers. A 0-1 value calculated using the averaged Mander's coefficients (plotted on the y-axis) shows that Hemotin-GFP has low overlap with Mitochondrial marker GRP75 (L3 0.13±0.02, n=18; WPP stage (0.31±0.02, n=14). There is high overlap of the Hemotin-GFP SEP with ER marker KDEL, particularly at the L3 stage (0.87±0.03, n=16) and lowered at the WPP stage (0.59±0.02, n=15). Overlap with Lysotracker is also very high at L3 (0.73±0.02, n=16) and increases at the WPP stage (0.85±0.01, n=15). Scale bars = 5μm.


A
GRP75
Hemotin GFP
DAPI
Merge
HmlGAL4>UAS-hemoGFP L3
B
GRP75
Hemotin GFP
DAPI
Merge
HmlGAL4>UAS-hemoGFP WPP
C
KDEL
Hemotin GFP
Merge
HmlGAL4>UAS-hemoGFP L3
D
KDEL
Hemotin GFP
Merge
HmlGAL4>UAS-hemoGFP WPP


Lysotracker-RED
Hemotin GFP
Merge
E
HmlGAL4>UAS-hemoGFP L3
Lysotracker-RED
Hemotin GFP
Merge
F
HmlGAL4>UAS-hemoGFP WPP


G
Co-localisation of hemo-smORF-GFP to Organelle markers
Mander's Coefficient
1.0
0.8
0.6
0.4
0.2
0.0
GRP75 L3
GRP75 WPP
KDEL L3
KDEL WPP
Lysotracker-RED L3
Lysotracker-RED WPP

quantified using the image analysis software ImageJ, with the Colocalisation Analysis plugin package (http://www.uhnresearch.ca/facilities/wcif). Mander's coefficient is a type of Pearson's correlation coefficient that can be used to define the covariance between signals from fluorophores in different channels (i.e. GFP and Rhodamine) using a numerical range from 0 (low colocalisation) to 1 (high colocalisation). The Mander's coefficient value for colocalisation between Hemotin-GFP and GRP75 at L3 (0.13±0.02) is quite low, however this overlap increases at WPP stage (0.31±0.02). Though the overall overlap between mitochondria and Hemotin may be low, the increase at this stage might indicate that there is a stronger interaction at WPP, possibly due to the increase in expression of Hemotin-GFP at pupal stages. Mander's coefficient for Hemotin-GFP and KDEL antibody in L3 haemocytes is very high (0.87±0.03) and is lowered in WPP haemocytes (0.59±0.02, n=15). Hemotin-GFP and Lysotracker, which labels organelles with low pH, give a Mander's coefficient value of 0.73±0.02 and the overlap is even greater at WPP stage, 0.85±0.01, as summarised in Figure 3.6G.

**Fly viability after wounding and infection**

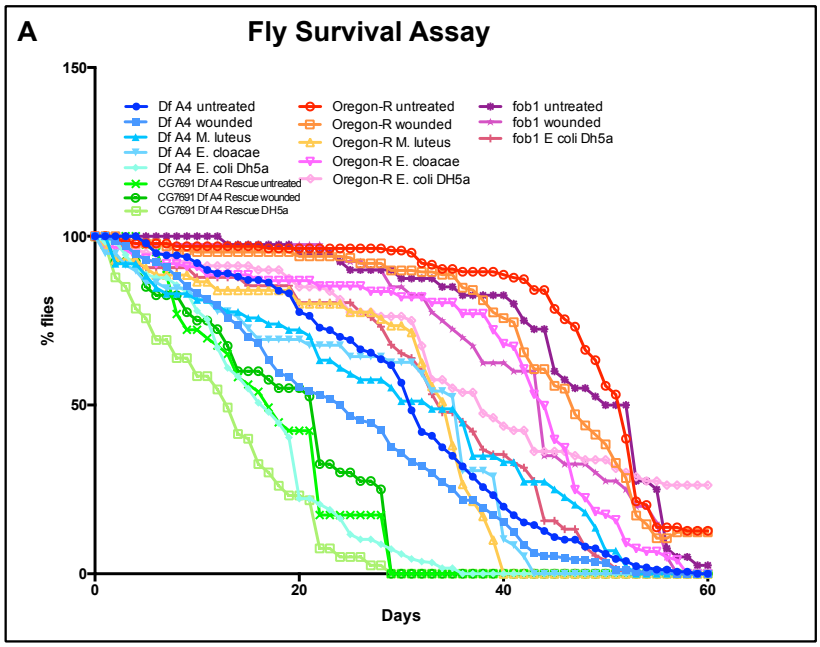
Due to the mounting evidence of the involvement of the Hemotin peptide in lysosomal and endosomal compartments, we decided to test the impact of infection and wounding on *Hemotin* mutant flies. The aims of this experiment was to observe whether the overall viability of *Hemotin* mutant flies was affected by lack of *hemotin* expression. We also wanted to see whether deleting the *hemotin* gene would be exacerbated by infection with various pathogens, thereby implicating a role for Hemotin in immunity via intracellular trafficking.

*Drosophila* display a differential induction of various AMPs after infection, and there is a possibility that Hemotin could be involved in the immune response in this manner, and is actuallt being secreted from the cell (via the ER and Golgi as shown by anti-KDEL co-localisation). This would become apparent in the flies infected with pathogenic bacteria if there is an impairment in the induction of the immune response. If Hemotin is mainly involved in the capacity of intracellular trafficking, this could potentially become apparent by infecting the flies with a concentrated dose of non-pathogenic *E. coli* (DH5α), similar to an experiment by Akbar *et al.* (Akbar *et al.* 2011) which showed

a requirement for the endocytic protein called full-of-bacteria (fob) in order to clear non-pathogenic bacterial infection. Mutated *fob* (*fob1*), flies become 'hypersensitive' to non-pathogenic bacterial infections and die earlier than *wild-type* flies, and we used this line as a positive control.

In order to verify the specificity of decreased viability to the *Hemotin* gene, I used a fly line that rescues the deletion of CG7691 in a *Df A4* background. Virgin males and females, 0-2 days old, were collected and kept at a maximum density of 20 flies per vial. For each strain, we kept three classes: untreated, wounded and infected. Infection was conducted by pricking the flies in the abdomen, just under the haltere, with a tungsten needle dipped into a concentrated culture of the Gram-positive, pathogenic bacteria *Micrococcus luteus*, the Gram-negative pathogenic bacteria *Enterobacter cloacae*, or a concentrated dose of the non-pathogenic DH5α *Escherichia coli*. As a control, flies of each strain were wounded in the abdomen with a tungsten needle dipped in sterile PBS. The flies that remained alive were then counted over the next 60 days (Figure 3.7A). I investigated the overall life-span viability of *Hemotin* mutant flies (*Df A4*) as compared to *Oregon-R (Or-R),* a commonly used *wild-type* strain, and *fob1* flies. This was compared to the result of non-pathogenic bacterial infections in these same lines.

The results of this experiment show that the overall lifespan of *Df A4* and the *Df A4;CG7691* rescue fly lines is significantly shorter when compared to both *Or-R* and *fob1* flies that undergo the same treatment. These results can also be depicted in a fashion similar to LD50, which is defined as the 'median lethal dose, 50%' of a given toxin or pathogen. As such, the table in Figure 3.7B depicts the average day upon which 50% of the flies are dead after each respective treatment or infection. Compared to *Or-R* flies (LD50 = 52.8±4.11 days), *Df A4* flies have a drastically reduced viability overall, without any infection or treatment (LD50 = 32.57±5 days), which is a 38% reduction from the LD50 of *Or-R* flies. Untreated *fob1* mutant flies (LD50 = 52.5±4.9 days), have comparable viability to *Or-R* flies. Gram-positive *Micrococcus luteus*, and Gram-negative *Enterobacter cloacae* were grown overnight from concentrated cultures, and used to infect the flies with a tungsten needle. Infection with pathogenic gram positive *M. luteus,* drastically decreased the viability of *Or-R* flies (LD50 = 35± 0 days), which




**B** **LD50 of fly viability**

| | control | wounded | *E. coli Dh5a* | *M. luteus* | *E. cloacae* |
|---|---|---|---|---|---|
| Oregon R | 52.75 ± 4.11 | 48.25 ± 4.86 | 40.5 ± 3.54 | 35 ± 0 | 43.5 ± 2.12 |
| Df A4 | 32.57 ± 5.0 | 25.29 ± 7.43 | 17.33 ± 4.61 | 28.0 ± 8.49 | 34 ± 1.41 |
| CG7691; Df A4 Rescue | 24 ± 2.83 | 21 ± 9.90 | 13.5 ± 0.71 | - | - |
| fob1 | 52.5 ± 4.95 | 47 ± 8.49 | 36 ± 2.83 | - | - |

**Figure 3.7: Fly viability after wounding and infection**

**A)** Fly viability assay showing percentage of flies that remain alive (y-axis) over a timespan of 60 days. Flies were either untreated, wounded as a control or infected with pathogenic or non-pathogenic bacteria in order to investigate the role of Hemotin in immunity. *Oregon R* flies were used as a negative control while *fob1* mutant flies were used as a positive control for DH5α infection. Overall, the *Df A4* line and genomic-rescue *CG7691, Df A4* flies have the lowest viability after all treatments. **B)** LD50 averages from all repeats of fly viability studies showing the number of days at which 50% of the flies are dead. *Df A4* flies have a shorter median life span compared to untreated *wild-type Oregon-R* flies across all treatments.

is 34% less than the untreated control. *M. luteus* infection also has large effect on the already less viable *Df A4* flies (LD50 = 28± 8.5 days), which is 20% less than *M. luteus* infected *Or-R* flies, and 15% less than *Df A4* overall median longevity.

Infection with pathogenic gram negative *E. cloacae* bacteria decreases the viability of *E. cloacae* infected *Or-R* (LD50 = 43.5±2.1 days) by 18%, compared to untreated *Or-R*. *E. cloacae* infected *Df A4* flies (LD50 = 34±1.4 days) do not show a significant effect by Gram negative infection when compared to untreated *Df A4* flies, in fact they live for 2 more days, on average. These results imply that the effect of pathogenic bacteria impacts both *wild-type* and *Hemotin* mutant flies in the same fashion, by decreasing their overall viability. We know from previous RT-PCR (MA Cespedes, personal communication) that the *Df A4* flies show an increase of *Diptericin* gene expression after infection similar to *wild-type* flies (not shown). Diptericin is an anti-microbial peptide involved in the Gram-negative infection response (Lee *et al.* 2001). There does not seem to be an increased lethality response upon pathogenic infection in *Hemotin* mutant flies.

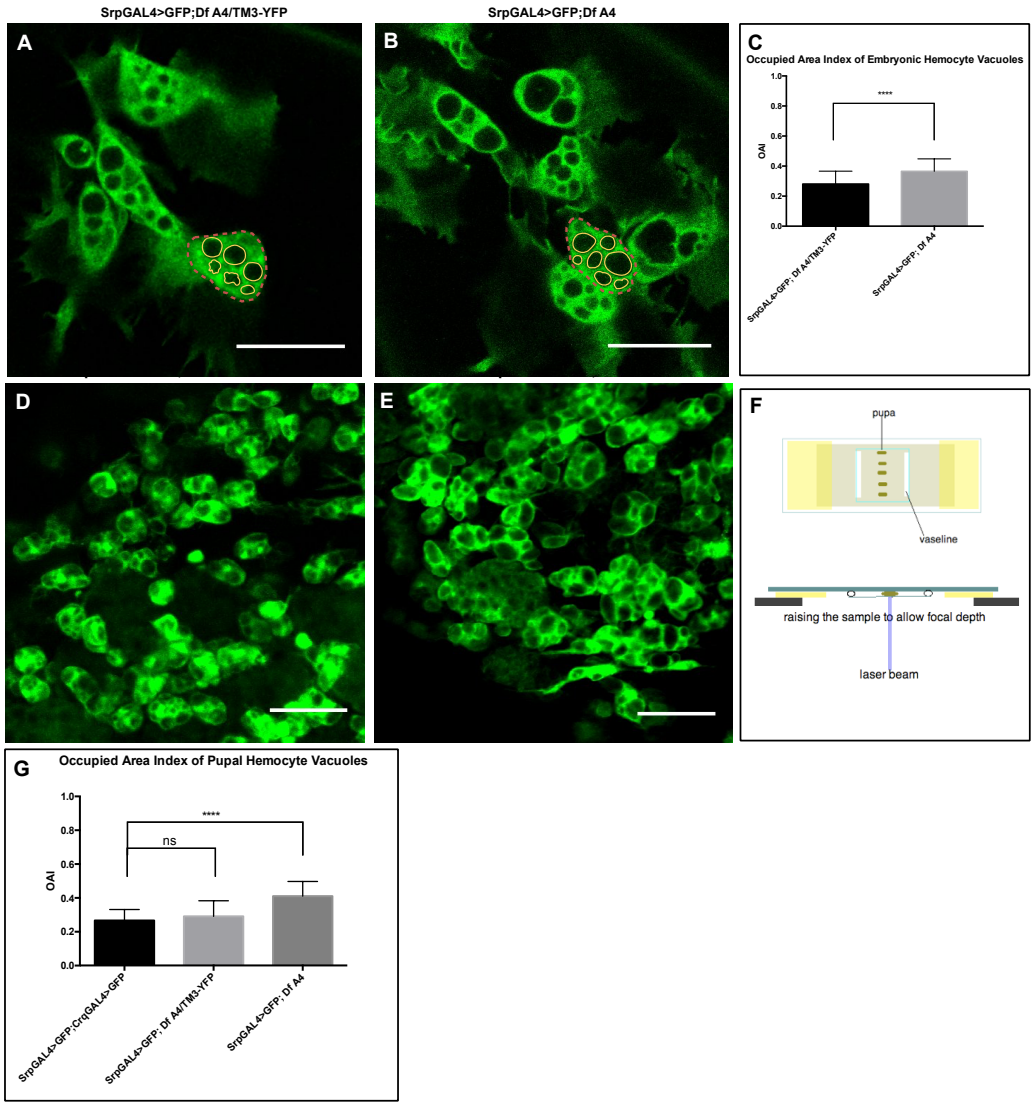
The most interesting results come from flies infected with the non-pathogenic *E. coli* DH5α which was performed in order to determine if Hemotin plays a role in intracellular trafficking. *Wild-type* flies have been shown to clear themselves of infection with non-pathogenic bacteria and survive, while mutants in the phagocytic pathway show an overall lowered viability because they are unable to digest large loads of bacteria (Akbar *et al.* 2011). The phagocytosis defect indirectly affects longevity via the phagocytic ability of haemocytes. For this experiment, a much more concentrated dose (100x) of non-pathogenic *E. coli* was used to infect the flies as compared to the concentrations of the pathogenic *M. luteus* and *E. cloacae* infections. The results from this study show a 23% decline in median life span of *Or-R* flies (LD50 = 40.5±3.5 days). DH5α infection of *Df A4* flies caused a 67% decrease in viability compared to untreated *Or-R*. There is a 47% drop in viability of infected *Df A4* flies, compared to untreated *Df A4*. The *fob1* mutant allele, which is described to have an endocytic/ phagocytic defect, is only marginally decreased in viability from DH5α infected *Or-R*, by 11%. By comparison, *Df A4* flies are 51% less viable than *fob1* mutant flies after DH5α infection. Infection with pathogenic and non-pathogenic bacteria in *Hemotin*

mutant flies dramatically decreases the viability of these flies. These results indicate that Hemotin may initially function in an intracellular capacity, which then later impinges on the overall immunity of the fly, as the haemocytes may be unable to clear an infection. This is similar to the story of the *full-of-bacteria* gene, which functions in phagocytosis and impinges on the overall health of the fly. Though *wild-type* animals also show a decrease in viability after infection, the viability of *Hemotin* mutant flies drops drastically, even when compared to the *fob1* positive control. This result also strongly suggested that Hemotin may be involved in intracellular trafficking and phagocytosis.

### *In vivo* imaging of haemocytes

If Hemotin is involved in the uptake and processing of material by haemocytes, one way to detect a possible phenotype would be to observe the behaviour and morphology of haemocytes at stages when their phagocytic functions are required. The availability of haemocyte-specific GAL4's, such as Serpent (Srp) and Peroxidasin (Pxn), which are expressed in the cytoplasm of haemocytes, allowed us to express UAS-GFP in these cells and visualise them at various stages of the *Drosophila* life cycle. By crossing flies containing these haemocyte-specific GAL4s driving the expression of GFP to our *Df A4/TM3-YFP* line, we were able to obtain *Hemotin* deficient flies with fluorescent haemocytes.

During *Drosophila* embryogenesis, haemocytes migrate from the presumptive head region, along the ventral midline to populate the entire embryo between stages 10-17 of development (Evans *et al.* 2010a). This method has been used to describe several types of embryonic haemocyte behaviours such as motility, wound recruitment and bacterial infection. It became apparent through Time-lapse imaging of embryonic haemocytes that *Df A4* embryonic haemocytes follow the described pattern of dispersion at embryonic stages, and hence there does not seem to be a haematopoietic defect at embryonic stages (I. R. Evans, personal communication). Using the method described by Evans *et al.* (Evans *et al.* 2010a) I was able to conduct imaging of haemocytes in live embryos during late-embryonic stages. The ability to visualise haemocytes *in vivo* allowed us to determine the effects of the *Hemotin* mutation at the cellular level.




**<u>Figure 3.8: Measuring vacuolar OAI in embryonic and pupal haemocytes *in vivo*</u>**

Haemocyte vacuoles were measured in haemocytes labelled with GFP in **A)** Hemotin mutant heterozygote (*Df A4/TM3-YFP*) embryos containing one copy of the *hemotin* gene. **B)** Hemotin null mutant (*Df A4)* embryos. The Occupied Area Index (OAI) was calculated by dividing the Vacuole area (solid yellow ROI) by the total cell area (dashed red ROI). Scale bar = 10µm **C)** the OAI is significantly larger in *hemotin* null embryonic haemocytes ($P<0.0001$). Haemocyte vacuoles measured in **D)** mutant heterozygote (*Df A4/TM3-YFP*) and **E)** Hemotin null mutant (*Df A4)* pupae. Scale bar = 10µm **F)** Pupal haemocyte imaging *in vivo* was achieved by cutting a window into the pupal casing and raising the sample by one slide thickness to increase focal depth. **G)** OAI averages of *hemotin* null *Df A4* pupae are significantly different ($P<0.0001$) to wild type and heterozygote pupal haemocytes.

In *w;srpGAL4>GFP;Df A4* homozygote embryos, haemocyte vacuoles are larger and slightly abnormal compared to the heterozygote condition (*w;srpGAL4>GFP; Df A4/ TM3-YFP*) (Figure 3.8A,B). As previously mentioned, the appearance of large cellular inclusion bodies, or vacuoles, are an indicator of pro-haemocytes becoming phagocytic at embryonic stages (Lanot *et al.* 2001). Measurements of the entire area of the haemocyte, as well as the vacuolar area within the cell, were obtained by taking Z stack images of live homozygous *Df A4* and heterozygous *TM3-YFP* embryos (red and yellow dashed lines, Figure 3.8A,B). From these measurements, an Occupied Area Index (OAI) was calculated by dividing the sum of total vacuolated area of a cell by the overall area of the cell; this ratio gives the total area of a haemocyte that is vacuolated. *Df A4* haemocytes (0.38) at late embryogenesis have significantly ($P<0.0001$) larger average OAI than *wild-type* embryos (0.29) (Figure 3.8C).

Fly lines *w;SrpGAL4>GFP; CrqGAL4>GFP* and *w;SrpGAL4>GFP; Df A4* were used to conduct live-imaging of pupal haemocytes. Stage P6 pupae (approx. 50h APF) were used, as this time point is well after the degradation of larval tissue and subsequent phagocytosis by the haemocytes released at pupation. Fly pupae are thick samples that present a challenge in laser confocal microscopy, usually restricting imaging to the superficial layers only (Nwaneshiudu *et al.* 2012). In order to get around this, pupae were mounted on double-sided tape, and the distance between the lens and the sample was increased by adding a 1-slide thick layer to the slide, enabling imaging of several cell diameters into the animal (Figure 3.8F). Opening a window in the pupal cuticle, over the thorax, allowed clearer imaging of fluorescent GFP-expressing haemocytes under the epidermis of the metamorphosing fly. (Figure 3.8D,E). The OAI for pupal haemocyte vacuoles in the thorax show that *Df A4* pupal haemocytes have larger vacuoles, or cellular inclusions, when compared to *wild-type* flies (Figure 3.8G). At this developmentally relevant time point, when haemocyte activity is necessary, *Hemotin* mutant haemocytes appear to contain a larger amount of phagocytised material. This supports the hypothesis that Hemotin could be playing a role in the regulation of phagocytosis or the breaking down of phagocytised material in haemocytes. Without the *Hemotin* gene, this process is hampered, giving rise to larger vacuoles in *Df A4* haemocytes, and a quantifiable phenotype of Hemotin peptide function.

**Labelling haemocyte vacuoles**

So far, the results of my experiments had suggested that Hemotin is involved in the processing of intracellular vesicles (or vacuoles), the peptide localises to Lysotracker and intracellular trafficking organelles, and is involved in immunity via phagocytosis. The main function of haemocytes is phagocytosis of extracellular material. Phagocytised material is brought into the cell by encapsulation by the plasma membrane. The internalised membrane-bound organelles are fused to organelles containing acidifying enzymes, at which point they are called lysosomes, the cytoplasmic sites where ingested particles undergo acid-mediated degradation. The membrane components of the organelles are then recycled to the plasma membrane edge (Velichkova *et al.* 2010; Jean *et al.* 2012). If Hemotin is involved in the processing of phagocytised material, we would potentially observe enlarged lysosomes, or lysosome-precursor organelles, in the cytoplasm of *Df A4* fly haemocytes (Luzio *et al.* 2000).

Using Pxn-GAL4 to drive the expression of UAS-GFP reveals cytoplasmic and nuclear signal of GFP in haemocytes *ex vivo*. Peroxidasin (Pxn) is an extracellular matrix protein that is specifically expressed in the haemocyte cytoplasm at all stages of *Drosophila* development (Nelson *et al.* 1994), allowing us to fluorescently visualise the haemocytes at any stage. The GFP-expressing haemocytes were extracted from *wild-type* and *Df A4* larvae from the fly lines *w;PxnGAL4>GFP; CrqGAL4>GFP* and *w; PxnGAL4>GFP; A4/TM3-YFP,* respectively, and cultured them on collagen-coated coverslips in growth media (DHIM). LysoTracker dye freely crosses cell membranes and fluoresces upon uptake into compartments with low pH, thus labelling lysosomes. After allowing the haemocytes to adhere to the collagen substrate, Lysotracker-Red dye was added to the growth media, and the haemocytes were live-imaged to reveal lysosomes in a cytoplasmic background labelled with GFP (Figure 3.9A,B). The LysoTracker staining in *ex vivo* primary haemocytes shows that the vacuoles in the cytoplasm of the *PxnGAL4>UASGFP* haemocytes are Lysotracker-positive in both *wild-type* and *Df A4* haemocytes. LysoTracker area was determined for each cell and frame by separating the Red (C1) and Green (C2) channels and using the 'Analyse particles' tool in ImageJ to determine the vacuole area (C1) and the entire area of the

cells(C2). Thirty-two cells were imaged over three technical repeats, with six larvae per repeat. The average OAI of LysoTracker area/Cell Area (C1/C2) is shown in Figure 3.9C. The average OAI of LysoTracker in *Df A4* haemocytes is significantly higher than that in *wild-type* haemocytes. This means that Hemotin mutants contain larger lysosomes than their wild-type counterparts, indicating a defect in the growth or processing of these intracellular vesicles.

Lysotracker stained WPP *ex vivo* haemocytes were also visualised over time, at 5-minute intervals, for a total of 150 minutes (Figure 3.9D) for a minimum of 15 cells per sample, and two technical repeats. This pulse-chase assay in WPP haemocytes was used to observe lysosome degradation in the cells over time, providing some insight as to whether *Df A4* haemocytes have impaired phagocytic and degradation capability compared to *wild-type* haemocytes. Although the overall difference in the OAI is quite low, when observed over time, the LysoTracker signal in *wild-type* haemocytes shows a cyclical pattern, with three distinct peaks approximately every 40-45 minutes (circles on data plot, Figure 3.9D). *Df A4* haemocytes are occupied by a larger lysosomal area from the start of the experiment, and there was a small initial peak at 20 minutes and a second one at 115 minutes (95 minutes later), which is approximately double the time it takes *wild-type* cells to process their lysosomes (squares on data plot, Figure 3.9D). These results showed that not only do *Df A4* haemocytes have a larger area occupied by lysosomes, but also that the lysosomal mass persisted over time. This could mean that Hemotin mutant haemocytes are ingesting more material from their environment, as well as having trouble processing the material that they have ingested. This extends the ‘larger vacuole’ phenotype observed in embryos and pupae to L3 and WPP lysosomes.

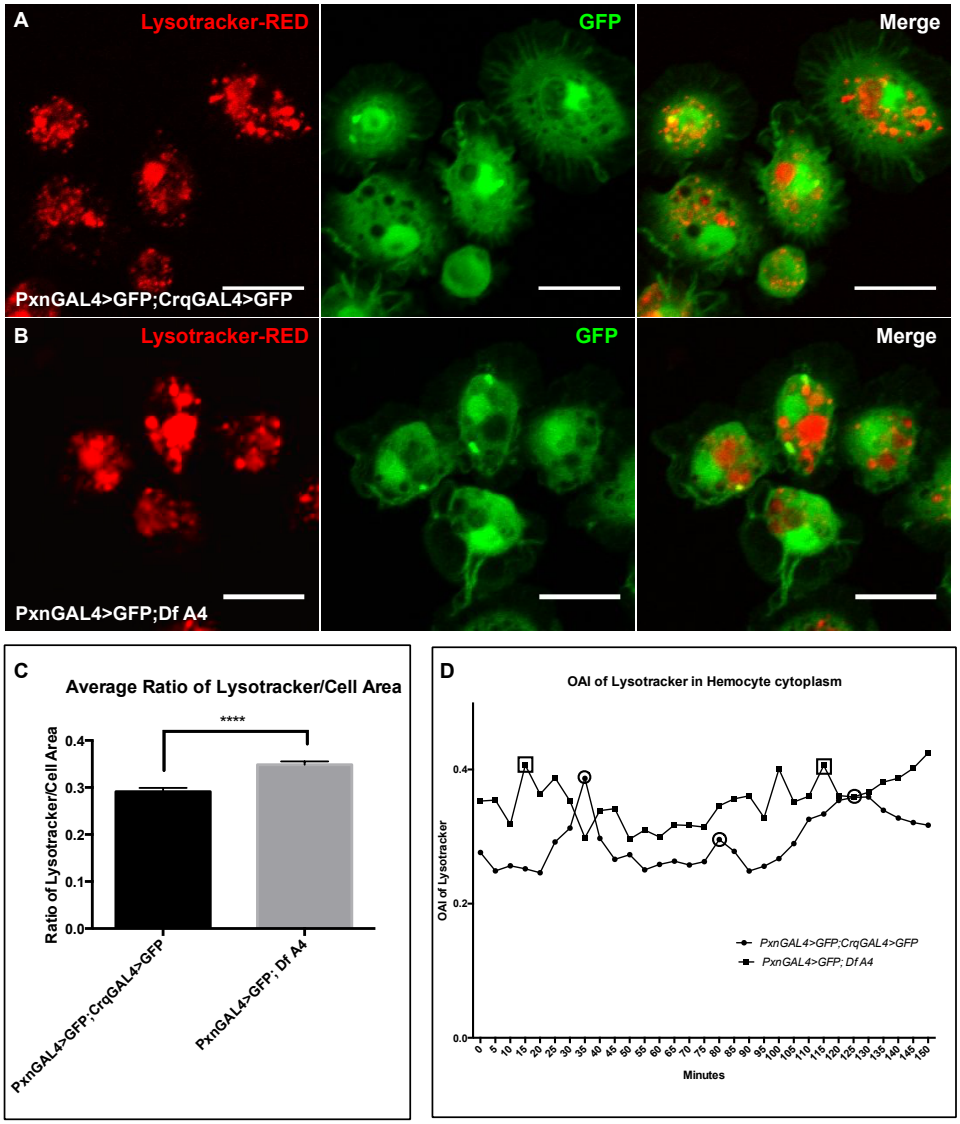



**Figure 3.9: Lysotracker-labelling of *ex vivo* L3 and WPP haemocyte vacuoles**

**A)** Haemocytes from *w;PxnGAL4>GFP; CrqGAL4>GFP* L3 animals show labelling of cytoplasmic vacuoles with Lysotracker-Red **B)** Lysotracker labelling of haemocytes from *w;PxnGAL4>GFP; Df A4 hemotin* null L3 animals. **C)** This allows us to calculate a lysosomal OAI which is significantly larger for Hemotin deficient L3 hemocytes ($P<0.0001$). **D)** Graph plotting the average OAI of Lysotracker (y axis) in WPP haemocytes over a 150 minute period. Data-points are at 5-minute intervals from the addition of lysotracker to the cells. Squares represent peaks of lysosomal area in *Df A4* haemocytes and circles depict peaks of lysosomal area in *wild type* haemocytes. Scale bars = 10μm

**Characterisation of haemocyte morphology *ex vivo***

Using the *ex vivo* primary cell culture technique described by Sampson *et al.* (Sampson 2013), a comparison of morphological classes of haemocytes extracted at L3 and WPP was conducted. In *wild-type* flies there is an observable 'switch' of haemocyte behaviour and morphology between larval and white pre-pupal (WPP) stages, from a largely sessile population (minor, normal cell types) to a pre-motile (non-polarised) or motile (polarised) population, a switch that we can now attribute to the 'activation' of haemocytes by ecdysone signalling (Sampson and Williams 2012; Sampson *et al.* 2013). Membrane turnover is an important part of the migratory and plasma-membrane activity of haemocytes. It has also been documented that haemocytes in embryos display a 'freezing' phenotype upon phagocytosis of *E. coli* coated with Mcf1 toxin, which they are unable to process and causes a halting of membrane dynamics (Vlisidou *et al.* 2009). Membrane turnover is therefore affected by phagocytosis of bacteria.

Using this established system of classifying haemocyte morpho-behaviour *ex vivo,* we characterised the membrane characteristics of *Df A4* haemocytes. If Hemotin is involved in the regulation of the cytoskeleton in haemocytes, this would become evident using this technique, and give us another avenue of investigation. A lack of normal morpho-behavioral characteristics *Df A4* haemocytes would also be consistent with the presence of enlarged vacuoles and lysosomes in the cytoplasm of the cells.

The classification system used in Sampson *et al.* (2012), which employs 6 categories of haemocyte spreading types, found in L1, L2, L3 and WPP stages, was simplified to suit the cell types found at L3 and WPP. Morphological classes were determined by scoring phase-contrast time-lapse videos of the *ex vivo* haemocytes that were bled from L3 or WPP flies of each genotype and then deposited on a collagen-coated coverslip inside a culture dish. The 'normal' (N) and 'minor' (M) haemocyte types, have a relatively uniform spread of the plasma membrane around the centroid regions (first and second rows, Figure 3.10) where the N haemocytes have a wider spread and M haemocytes have a tighter spread of membrane. Both N and M spread represent the quiescent, sessile behaviour and morphology of haemocytes before they are activated by ecdysone and apoptosis. These cells remain relatively unchanged in regards to their morphology and membrane behaviour over the course of a 20 minute time-lapse. The two

morphological classes of haemocytes that are predominantly found at WPP stage are 'non-polarised' (NP) and 'polarised' (P) types (third and fourth rows, Figure 3.10). Both classes represent a fully adhered cell with an active cell membrane that protrudes and retracts from the centroid region, resulting in membrane ruffling. The essential difference between these two classes is that P cells produce a leading edge (solid arrowheads, Figure 3.10), and display directional migration, while NP cells throw out protrusions in different directions (white arrowheads, Figure 3.10), without displaying directional movement. Anywhere between five and thirty cells can be observed in each 20-minute time lapse video. In order to make the haemocyte numbers more quantitatively viable, the total cell numbers were added together to give a more complete profile of the haemocyte population. A minimum of fifty cells per genotype were collected over two technical repeats from a sample pool of mixed bleeds from six larvae or six WPP (summarised in Figure 3.11).

For the L3 stage *ex vivo* characterisation both *Or-R* and *w1118* flies were used as *wild-type* controls, and haemocytes from both lines displayed a mixture of M, N, and NP cell types, with *Or-R* L3 haemocytes also containing a small percentage of cells with directional motility (P). *Hemolectin*(*Hml) GAL4 (III)* flies were crossed with a *UAS-GFP* line in order to determine if the ectopic expression of GFP by various haemocyte specific GAL4's has an effect on haemocyte cellular profiles. These profiles were compared to the previously described fluorescent haemocyte lines, which served as *wild-type* controls for vacuole measurements, and to see whether there is a deviation from non-GFP expressing haemocytes bled from *Or-R* and *w1118* flies. The expression of GFP with a haemocyte GAL4 driver shows similar profiles to the *wild-type* controls. At the L3 stage, it is apparent that *Hemotin* null mutant *(Df A4)* and the expression of a dsRNA designed to deplete the *Hemotin* mRNA (*HmlGAL4>UAS-Hemotin dsRNA*) decreased the overall percentage of P cells, with the M and N 'inactivated' types being predominant. Conversely, over-expression of full length *Hemotin* constructs showed *wild-type* distributions of cell types. With *HemeseGAL4,* which is a stronger driver than *HmlGAL4*, (observed by strength of GFP expression in haemocytes), several motile P cells, were also observed at L3. Notably, expression of full length *UAS-Hemotin*, *UAS-hemo-ORF1* and full-length *UAS-Hemotin-GFP,* expressed in a *Df A4* background,

rescued the effect of the null mutation at L3 by returning the haemocyte population to a *wild-type* state. *UAS-Hemotin-ORF1* is a mini-gene consisting of a fragment of mRNA truncated after the *hemotin-ORF1* stop codon. *UAS-Hemotin-ORF2* consists of an mRNA fragment carrying the ORF2 sequence and 6nt upstream of its start codon. These nucleotides are retained in order to conserve the endogenous Kozak sequence.(Pueyo 2016).

At WPP stage, the various control populations of haemocytes from *Or-R, w1118*, and GFP-expressing lines contained largely non-polarised or polarised cell types. *Df A4* heterozygote haemocytes are also similar to *wild-type* in their cell type profiles. There is a lingering presence of some Minor and Normal cells, but this tends to remain around 5-10% of the overall population. The striking observation in *Df A4* haemocytes is that they do not display the same proportion of motile P cells at WPP stage, an effect which can also be seen in the *Hemotin* RNAi line. WPP from the cross of *HmlGAL4* and *UAS-Hemotin-ORF1* and *UAS-Hemotin-ORF2,* showed an increase in NP cell types. This result is different to the haemocyte profile of *HemeseGAL4* driven ORF1 and ORF2 expression. This could be due to the fact that the peak of *Hml* expression occurs at larval stages, and may be actively repressed after the induction of pupation (Goto 2003). *HemeseGAL4* driven expression of UAS-Hemotin is closer to the haemocyte profile of *wild-type* WPP than with *HmlGAL4*.

*HmlGAL4* driven expression of full length *UAS-Hemotin* and *UAS-Hemotin-ORF1* in *Df A4* background, seems to partially rescue the *Df A4* phenotype of lowered number of P cells. I was unable to complete the rescues using *HemeseGAL4,* which may have shown an improved effect of the rescue using the Hemotin mini-gene constructs. The full length *UAS-Hemotin tagged with GFP* does not rescue the effects of *Df A4* at WPP stage. Results gathered by others at the same time showed that when *Df A4* L3 haemocytes were incubated with ecdysone, there was no induction of motility in those haemocytes (Christopher Sampson, personal communication). This is similar to the result of the dominant-negative mutants of ecdysone receptor at L3 stage (Sampson 2013). Overall, these results show that lowering or eliminating the expression of Hemotin by a specific deficiency, caused deviation from *wild-type* haemocyte profiles.

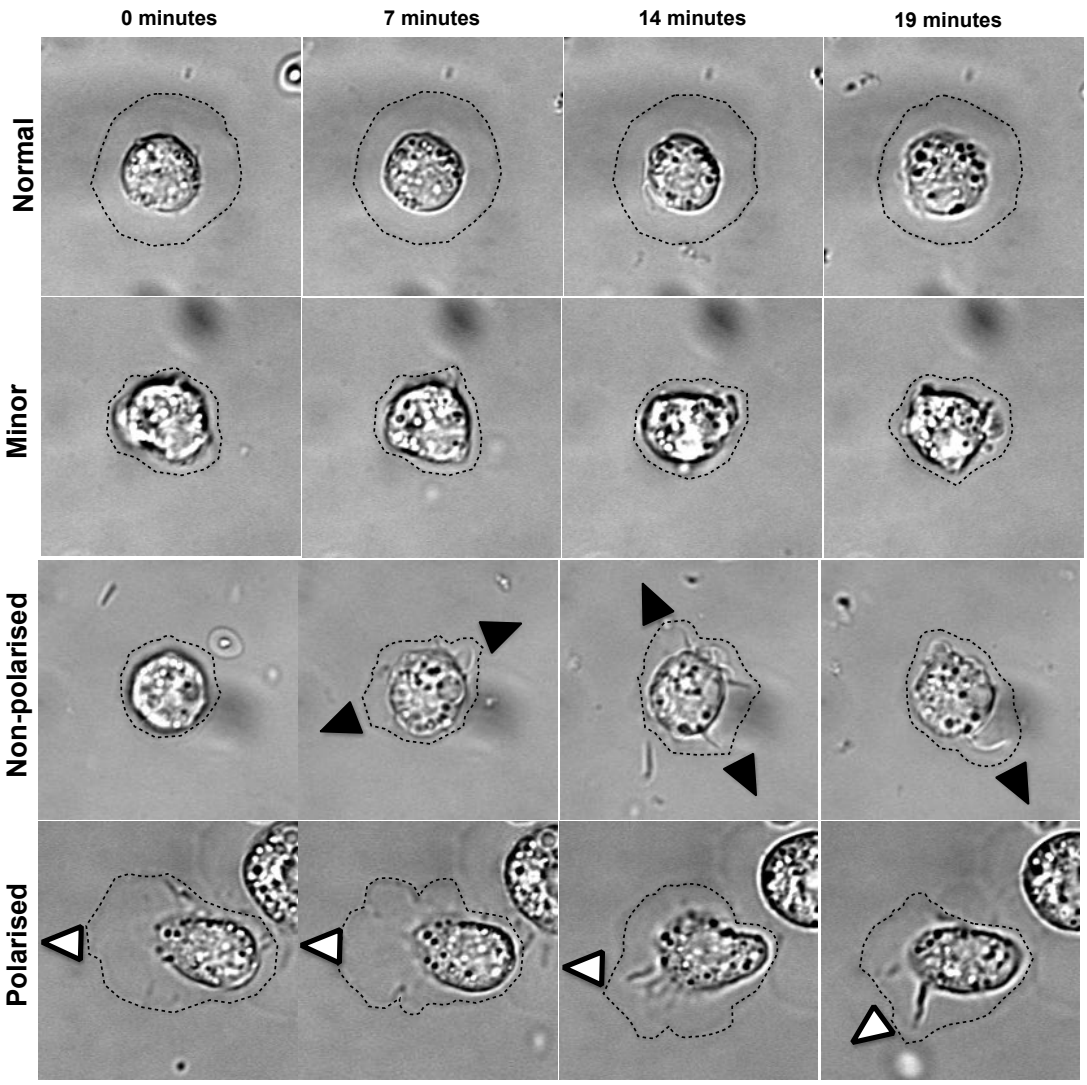
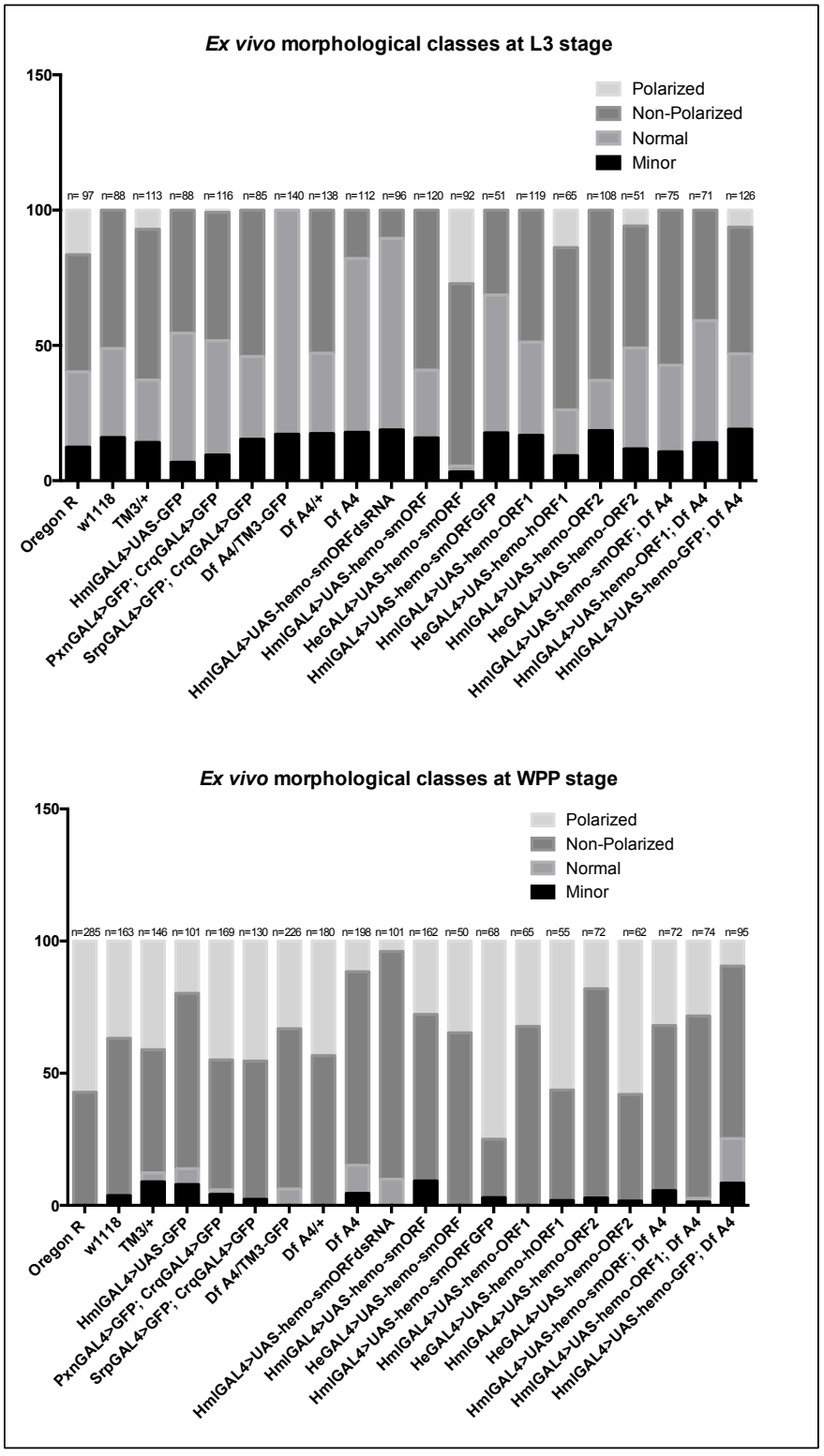



**Figure 3.10: *Ex vivo* morphological classes of L3 and WPP haemocytes**

Time-lapse imaging of *ex vivo* hemocytes and classification of cell behaviour through morphology over time. Minor (M) and Normal (N) spread classes are shown in the top two rows, and it can be seen that over the course of 19 minutes, the shape of these cells remain largely unchanged. These cell types are predominantly seen at the L3 stage. Polarised (P) and Non Polarised (NP) haemocytes show a dynamic cell edge which depict the more active haemocytes that can be predominantly seen in WPP stage haemocytes. The difference between these two classes is that NP haemocytes extend lamellipodia in different directions (black arrows) whilst P haemocytes display directional movement (white arrows).

**Figure 3.11: Results of e*x vivo* morphological classification of L3 and WPP haemocytes** Morphological classification of *ex vivo* hemocytes isolated from various controls expressing GFP in haemocytes, homozygote and heterozygote *hemotin* deficiency mutants, *hemotin* RNAi knockdown, and over-expression fly lines at Larval (Top graph) and WPP stages (Bottom graph). *Oregon R* and *w1118* L3 and WPP were used as controls to characterise the wild-type distribution of hemocytes. The profiles obtained for haemocytes extracted from transgenic animals at both these stages show that Hemotin depletion leads to a change from *wild-type* distributions, with a fewer proportion of activated (P, NP) haemocytes at both L3 and WPP stages in Hemotin mutants. Over-expression of Hemotin-ORF1 and Hemotin-full-length constructs results in an increase in activated hemocyte types (P, NP) at the L3 stage compared to wild-type haemocytes. The *Df A4* deficiency phenotype of haemocyte morphologies is partially specific to Hemotin as it can be partially rescued by the over-expression of the Hemotin-ORF1 and Hemotin-full-length constructs in a *hemotin* null background at both the L3 and WPP stages.


*Ex vivo* morphological classes at L3 stage
Polarized
Non-Polarized
Normal
Minor
150
100
50
0
n= 97 n=88 n=113 n=88 n=116 n=85 n=140 n=138 n=112 n=96 n=120 n=92 n=51 n=119 n=65 n=108 n=51 n=75 n=71 n=126
Oregon R
w1118
TM3/+
HmlGAL4>UAS-GFP
PxnGAL4>GFP; CrqGAL4>GFP
SrpGAL4>GFP; CrqGAL4>GFP
Df A4/TM3-GFP
Df A4/+
Df A4
HmlGAL4>UAS-hemo-smORFdsRNA
HmlGAL4>UAS-hemo-smORF
HeGAL4>UAS-hemo-smORF
HmlGAL4>UAS-hemo-smORFGFP
HmlGAL4>UAS-hemo-ORF1
HeGAL4>UAS-hemo-hORF1
HmlGAL4>UAS-hemo-ORF2
HeGAL4>UAS-hemo-ORF2
HmlGAL4>UAS-hemo-smORF; Df A4
HmlGAL4>UAS-hemo-ORF1; Df A4
HmlGAL4>UAS-hemo-GFP; Df A4
*Ex vivo* morphological classes at WPP stage
Polarized
Non-Polarized
Normal
Minor
150
100
50
0
n=285 n=163 n=146 n=101 n=169 n=130 n=226 n=180 n=198 n=101 n=162 n=50 n=68 n=65 n=55 n=72 n=62 n=72 n=74 n=95
Oregon R
w1118
TM3/+
HmlGAL4>UAS-GFP
PxnGAL4>GFP; CrqGAL4>GFP
SrpGAL4>GFP; CrqGAL4>GFP
Df A4/TM3-GFP
Df A4/+
Df A4
HmlGAL4>UAS-hemo-smORFdsRNA
HmlGAL4>UAS-hemo-smORF
HeGAL4>UAS-hemo-smORF
HmlGAL4>UAS-hemo-smORFGFP
HmlGAL4>UAS-hemo-ORF1
HeGAL4>UAS-hemo-hORF1
HmlGAL4>UAS-hemo-ORF2
HeGAL4>UAS-hemo-ORF2
HmlGAL4>UAS-hemo-smORF; Df A4
HmlGAL4>UAS-hemo-ORF1; Df A4
HmlGAL4>UAS-hemo-GFP; Df A4

**A putative vertebrate homolog of Hemotin**

I ran the peptide sequence of Hemotin through the Protein Homology/analogY Recognition Engine (Phyre2) program (Kelley *et al.* 2015) as it had been successfully used to find the homolog of Sarcolamban in vertebrates (Magny *et al.* 2013). It appeared that Hemotin (Figure 3.12A) could have a putative homolog in vertebrates called Stannin (Snn) (Figure 3.12B). There are a few early papers published on the characterisation and function of Snn, with experiments conducted mainly in mammalian tissues after the gene was isolated in toxicology studies (Toggas *et al.* 1992). Snn encodes an 88 amino acid peptide expressed in the rat spleen, brain, kidneys and lungs. Antibody staining against FLAG, a commonly used protein tag, fused to the Snn transcript, expressed in mammalian NIH-3T3 cells showed that Snn-FLAG fusion protein is expressed in a punctate pattern all over the cytoplasm of the cells, similar to the appearance of Hemotin-GFP expression in haemocytes (Figure 3.12C) (Davidson *et al.* 2004). They also tested whether there is overlap of Snn expression to the mitochondria, and they found that Snn-FLAG and GRP75 do colocalise (Figure 3.12D). Cell fractionation experiments followed by western blotting showed that Snn-FLAG was found to be enriched in peroxisomes and the ER (labeled with anti-GRP78 antibody) (Davidson *et al.* 2004; Billingsley *et al.* 2006). The Snn peptide sequence contains a CXC metal-binding motif, which is a conserved chemokine receptor, a putative 14-3-3ζ RxSxpSxD binding motif on the cytoplasmic side of the molecule and a 25aa transmembrane domain, which can be aligned to the amino acid residues in Hemotin (Billingsley *et al.* 2006) (Figure 3.12E). 14-3-3ζ has been shown to co-immunoprecipitate with Stannin (Davidson *et al.* 2005). Knocking down 14-3-3ζ affects phagocytosis in *Drosophila* haemocytes, as depletion of 14-3-3ζ leads to excessive actin polymerisation and lack of protrusion formation (Ulvila *et al.* 2011).

**Colocalisation of Hemotin-GFP and 14-3-3**

The availability of several specific anti-14-3-3 antibodies allowed the verification of the functional homology of Snn and Hemotin by colocalisation experiments. (Santa Cruz) was checked for colocalisation against Beta-tubulin (DSHB) as a positive control for the antibody, as 14-3-3 is known to localise to microtubules (Zhou *et al.* 2010). The 14-3-3

protein has several isoforms, and is a highly conserved master regulator of eukaryotic cells. Different isoforms of 14-3-3 are expressed in different cell types, and these peptides participate in neuronal function, regulation of cell cycle, apoptosis, exocytosis, as well as cytoskeletal dynamics and signalling (Fu *et al.* 2000). Using extracted haemocytes from *wild-type Or-R* L3 and WPP animals, the colocalisation of 14-3-3 and B-tubulin was confirmed (Figure 3.13A,B). I then proceeded to check whether Hemotin fused with GFP and 14-3-3 colocalisation could be observed. In a Z-stack projection of L3 and WPP haemocytes, there are spatial overlaps between the expression of Hemotin-GFP and 14-3-3 labelled domains (Figure 3.13C,D). Colocalisation of Hemotin-GFP to 14-3-3 is higher at L3 stage than at WPP. This shows that there is a moderate amount of overlap between Hemotin-GFP and 14-3-3, perhaps enough to mediate an interaction at specific foci. The average Mander's correlation coefficient between Hemotin-GFP and 14-3-3 (Figure 3.13E) was determined for L3 (0.5643 ± 0.03) and WPP cells (0.4678 ± 0.03). The Mander's coefficient for overlap between 14-3-3 and B-tubulin at L3 (0.8516 ± 0.02) is a little higher than at WPP (0.8106 ± 0.02). I have shown that Hemotin co-localises to similar intracellular membranes in haemocytes as Stannin does in vertebrate cells, and that Hemotin mutants have a defect in the processing of endocytosed material. The functional characterisation of the actual steps of endosome maturation that Hemotin is involved in were researched by other members of the lab, as will be discussed in the following section.

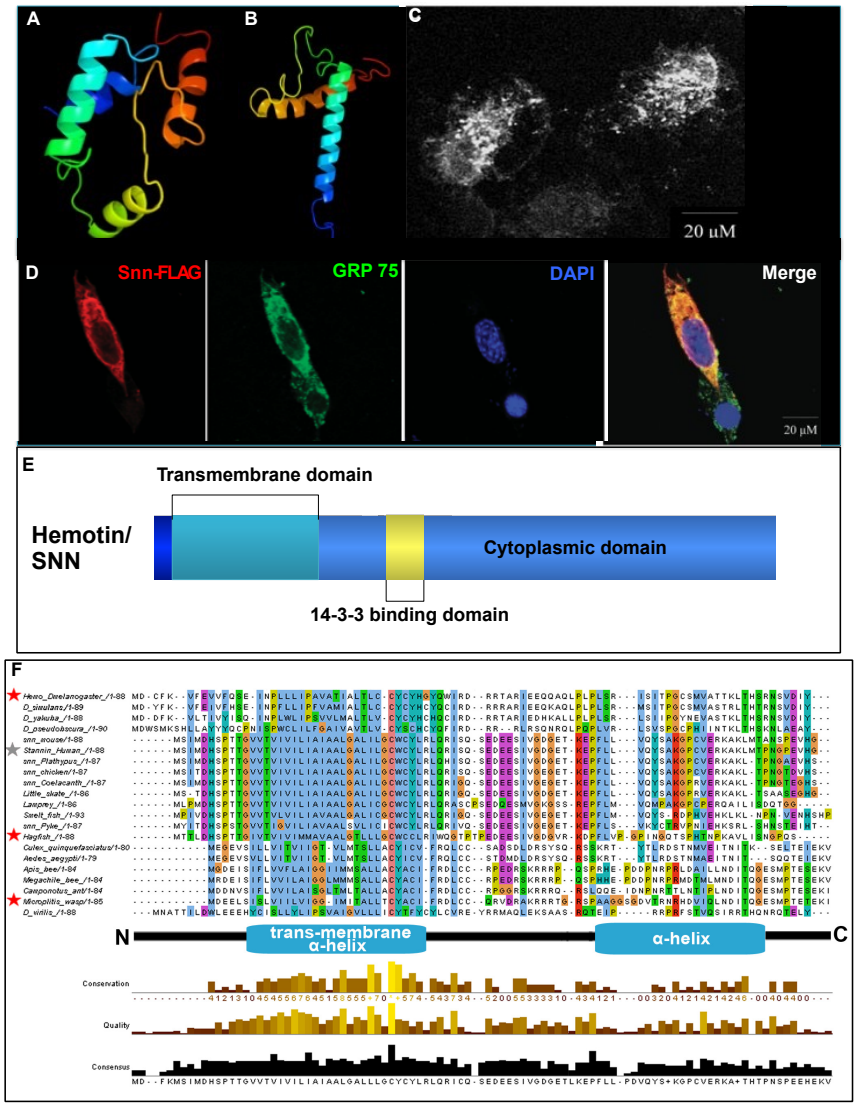



**Figure 3.12: The discovery of Stannin, a putative mammalian homologue of Hemotin A)** Phyre2 structural homology program predicted structure of Hemotin-ORF1 and **B)** The putative vertebrate 88 amino acid homologue of Hemotin-ORF1, called Stannin, as shown by the Phyre2 program **C)** Stannin-FLAG expression in mammalian NIH-3T3 cells shows a punctate subcellular distribution (Davidson *et al.* 2004) **D)** Colocalisation of Snn-FLAG with mitochondrial marker GRP75 shows a high level of overlap (Billingsley *et al.* 2006). **E)** Domain structure of the Stannin/Hemotin peptide (ORF1, 88amino acids) showing the putative transmembrane features and the 14-3-3ζ binding domain. **F)** Clustal Alignment of Hemotin shows the conservation of the peptide from (stars) two insect lineages (Diptera and Hymenoptera) and vertebrate Stannin (Hagfish and Human). There is a high level of conservation in the transmembrane domain motif among these peptides.

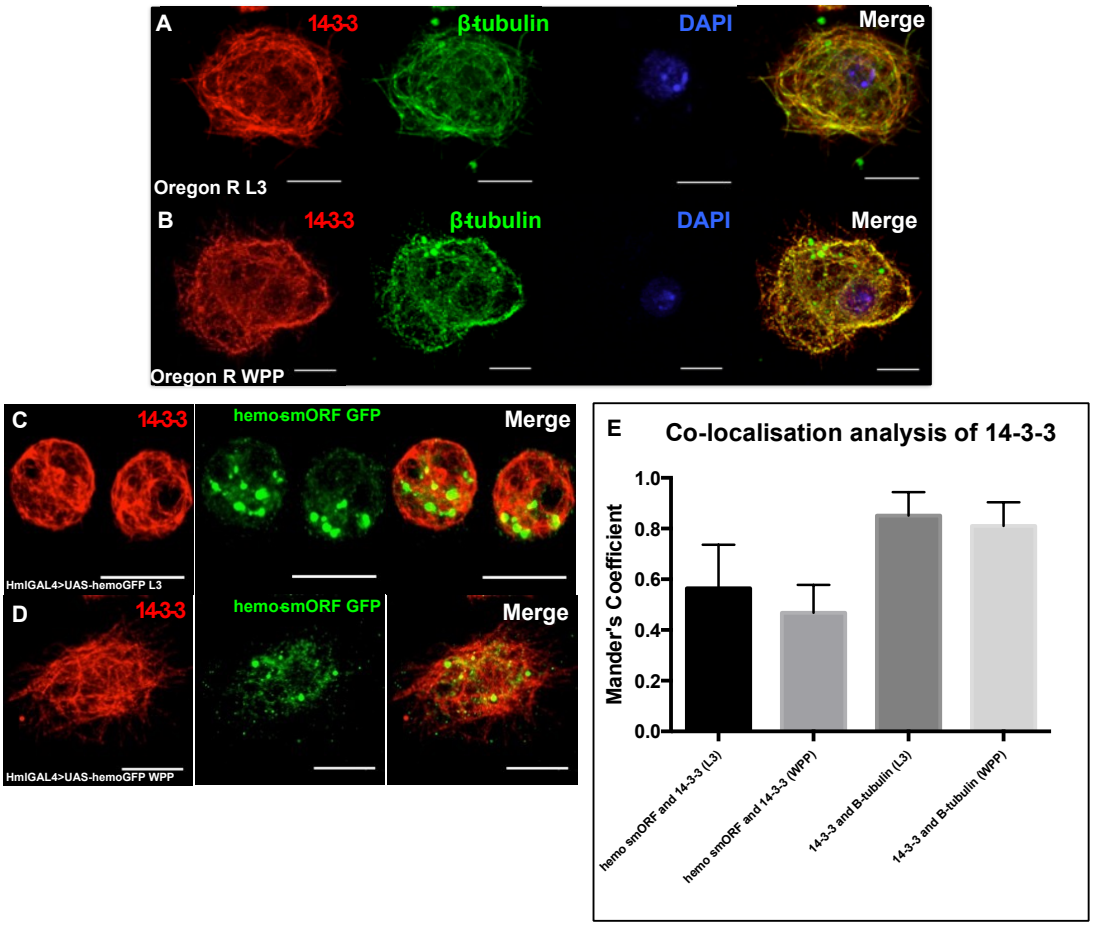



**Figure 3.13: Colocalisation of Hemotin-GFP and anti-14-3-3**

**A)** Double Antibody staining of *ex vivo* haemocytes to confirm colocalisation of a pan-isoform anti-14-3-3 antibody (Santa Cruz) (Rhodamine secondary) with anti-Beta-Tubulin (FITC secondary) and DAPI staining the nuclei (Blue) in *Oregon-R* L3 Larvae and **B)** WPP haemocytes. **C)** *HmlGAL4>UAS-Hemotin-GFP* haemocyte staining with pan isoform anti-14-3-3 antibody in haemocytes from L3 and **D)** WPP. Scale bars = 5µm. **E)** Quantification of colocalisation using Mander's coefficient. There is very high overlap between 14-3-3 antibodies and Beta-tubulin (L3=0.85±0.02, n=24; WPP= 0.47± 0.03, n=11) Hemotin-GFP and 14-3-3 peptides at L3 show medium level of overlap at L3 stage (0.56±0.03, n=25), and this overlap is decreased at the WPP stage (0.47± 0.03, n=11).

# Discussion

*Hemotin* is a dicistronic smORF gene, which produces two peptides of 88 and 59 amino acids from ORF1 and ORF2 respectively. We had previously found that *hemotin* is expressed in regions where haemocytes are derived from in embryos, similar to the expression of haemocyte-specific markers. *Hemotin* mRNA is not maternally deposited in the embryo, and the expression of *hemotin* peaks at the end of larval and beginning of pupal stages. The suspicion that *hemotin* mRNA is expressed in haemocytes at embryonic stages led us to investigate whether Hemotin played a role in *Drosophila* immunity, overall longevity of the fly, and the cellular process of intracellular trafficking. Haemocytes are excellent cellular models as they are well characterised cells. Haemocytes can be bled from the animal at post-embryonic life stages and survive for long periods in primary cell culture in *ex vivo* chambers supplemented with culture medium (Sampson and Williams 2012). The affinity of Haemocytes for extra-cellular matrix proteins such as collagen, enables live imaging on collagen-coated coverslips using organelle dyes and direct labelling of cells by fixing them and bathing the cells with antibodies (Sampson and Williams 2012; Sampson *et al.* 2013).

**Hemotin localises to KDEL and Lysotracker labelled organelles**

As we could not determine the endogenous expression of Hemotin peptide, in order to visualise Hemotin localisation with high resolution microscopy at the sub-cellular level, we had to express a UAS construct containing Hemotin tagged at the C-terminal with GFP, expressed with haemocyte-specific GAL4's. This method can also be used as proof of protein translation, as the GFP coding sequence would only be translated if there is translation of the *Hemotin* coding sequence. I showed that Hemotin-GFP colocalises with different membrane markers inside the cell, specifically Lysotracker, which labels acidic compartments in the cell; and KDEL which is a peptide motif presented on the trans-side of the ER, the cis-side of the Golgi, and any intracellular vesicles inside the cell. The presence of a large GFP tag (25KDa) on a small peptide could potentially distort the expression and subcellular localisation, but in the cases of both Tal and Scl, tagged versions of the smORF rescued lack-of-function phenotypes, showing that the fusion proteins were functional. In both previous cases, tagging the

smORF with a marker such as GST, GFP and FLAG was only way that smORF subcellular expression was detected.(Galindo *et al.* 2007; Magny *et al.* 2013). We decided to continue using haemocytes as a model as the effects of intracellular trafficking would become apparent in these cells.

**Hemotin plays a role in *Drosophila* immunity via intracellular trafficking**

By performing viability studies on the Hemotin-deficiency mutant (*Df A4*), it became apparent that the lack of Hemotin expression causes a decrease in the life span of flies. Although it became clear that Hemotin performs an important function in this way, we had yet to actually characterise exactly why the function of Hemotin would impinge upon overall life-span, as flies appeared normal until their early deaths. To discern whether Hemotin plays a role in immunity, flies were infected with pathogenic and non-pathogenic bacteria. This showed that Hemotin mutants are more susceptible to infection by pathogenic and non-pathogenic bacteria, which decreases their life spans significantly when compared to normal flies. Infected mutants of an endosome maturation protein called full-of-bacteria (Akbar *et al.* 2011) live longer than Hemotin mutant flies in this experiment. This led us to conclude that Hemotin is definitely involved in intracellular trafficking or endocytosis.

Overall fly viability is a measurable phenotype, but to prove the Hemotin functional hypothesis, required taking a closer look at haemocyte function. Embryonic and pupal stages are important times for haemocytic function, and imaging haemocytes *in vivo* at both stages showed that Hemotin mutant haemocytes have larger vacuoles than wild-type haemocytes, indicating a problem with phagocytosis or processing of ingested material. In addition, a LysoTracker pulse chase assay revealed larger lysosomal compartments in Hemotin mutant haemocytes *ex vivo,* after the start of metamorphosis. Hemotin mutant haemocytes also displayed a delay of lysosomal processing when comparing the measurements of cyclical increasing/decreasing lysosomal area to wild-type haemocytes. This further indicates that there is a defect in endosome maturation, as lysosomal contents are supposed to be degraded and through subsequent fusion events, the membrane that is brought inside the cell through phagocytosis is recycled to the cell edge (Lindmo and Stenmark 2006; Huotari and Helenius 2011; Akbar *et al.* 2011).

If the removal of Hemotin results in a defect in the endosome maturation pathway, this would potentially be reflected in the overall cytoskeletal morphology and the motile characteristics of Hemotin mutant haemocytes. By culturing the primary haemocytes using the *ex-vivo* system, I attempted to show that there are reduced characteristics of motile behaviour in *hemotin* mutant haemocytes at both L3 and WPP stage. As previously described, it is through the action of the Rho-family GTP-ases Cdc42, Rac1 and Rac2 causes the ruffling of plasma membrane, which allows haemocytes to probe their environment in order to recognise apoptotic debris and bacterial pathogens. This induces a receptor-ligand mediated signalling cascade which acts upon the actin scaffold of the cytoskeleton, allowing phagocytosis to occur. The switch in membrane behaviour from a largely sessile population to a motile, 'pathogen-seeking' haemocyte population does not take place in haemocytes derived from loss-of-function mutants of the Rho-family GTP-ases (Sampson and Williams 2012). I found that *hemotin* mutant haemocytes show a different set of cell morphological classes than wild-type cells, at both L3 and WPP stages. This phenotype could be a result of a phagocytic block inside the cytoplasm of cells, which hampers membrane turnover and leads to a block in the membrane ruffling and directional motility of haemocytes. These phenotypes were either completely (L3) or partially (WPP) rescued by expressing a full length construct of Hemotin in the haemocytes, returning the profile of the extracted haemocytes to more wild-type proportions.

This result, though not a direct read-out of Hemotin function, supports the idea that Hemotin is affecting membrane turnover, and was consistent with our hypothesis, considering that Hemotin mutants seem unable to process intracellular vesicles efficiently. Though, as it later turned out, the inhibition of motile characteristics can also be attributed to the deletion of *CG7691* in the *Df A4* line (C.J. Sampson, personal communication). This discovery helped explain some of the confusion around the *ex vivo* characterisation of *hemotin* mutant haemocytes, and also why rescues with the *UAS-hemotin* constructs only partially rescued the non-motile phenotypes. Direct comparison of the cell morphology profiles to that of actin cytoskeleton mutants, or phagocytosis mutants, would have made this analysis more reliable. The fact that ecdysone fails to stimulate *Df A4* haemocytes into activation also remains to be further

investigated.

Recently, we were able to confirm that Hemotin is indeed a novel regulator of the endosome maturation pathway. Other members in the lab have found that Hemotin localises to specific endosomal compartments which are labeled by markers of the early-to-mid transition of endosome maturation. The vacuoles I was previously measuring in *hemotin* mutants can be labeled with markers specific to late endosome maturation (Rab7, FYVE) and show a significant increase in size compared to wild type. This phenotype of enlarged late-endosomal compartments is fully rescued by the expression of *hemotin* in a null-mutant background. The knock−down of Stannin in a mammalian macrophage−like cell line also showed enlarged LysoTracker vacuoles, indicating a functional homology of Hemotin in vertebrates (Pueyo *et al.* 2016).

The polyclonal antibody that was generated against Hemotin did not exhibit specificity despite best efforts and different immunohistochemistry techniques employed. It would have been helpful as a tool to detect, visualise, and quantify endogenous Hemotin protein levels. As there was always strong, comparable levels of background signal present in *Df A4* mutant embryos and haemocytes, I was unable to show that the antibody specifically binds to Hemotin,. It has previously been verified by genomic and RT-PCR by others in our group that *Df A4* mutants completely lack *hemotin* at the genomic and RNA level, therefore it could not be possible that the antibody is specific to Hemotin. The non-specific antibody hampered our ability to look at a bigger picture of endogenous Hemotin expression, as there may be other tissues where Hemotin could be playing an important role such as the CNS, which is supported by the expression of *hemotin* in a CNS-derived cell line from modENCODE RNA expression data. Regardless of these shortcomings, finding a role for Hemotin in a context of phagocytosis and intracellular trafficking is very exciting.

By combining the knowledge of the localisation of Hemotin and using the deficiency mutant to functionally characterise a phenotype involved with the localisation of the peptide, I was able to find the first breadcrumbs on the trail to the functional characterisation of this novel SEP. This information helped others to study the role of Stannin in the context of endosome maturation, which had not previously been described. The information presented here may enlighten our understanding of

phagocytosis in higher animals as well as in *Drosophila*. The use of tagging and subcellular characterisation also drove us toward the broader eukaryotic concept of intracellular trafficking. Although the overall effect of stalling endosome maturation would not cause lethality in early life stages, the life span and quality of life of the flies are affected in *hemotin* deficiency mutants, especially at adult stages, which can be seen from the viability and infection studies.

# Chapter IV - Cellular characterisation of translated *Drosophila* smORFs

## Introduction

Characterising the smORF gene family on an individual basis is a time consuming process, requiring years of work from the discovery of a smORF gene, to the completion of functional characterisation at a molecular or cellular level. Previously, the characterisation of *tarsal-less*, *sarcolamban* and *hemotin* have taken up to 5 years each. The process of smORF characterisation on an experimental basis can only begin after the verification of a translated peptide product. Bioinformatic studies of sequenced genomes have revealed that hundreds and potentially thousands of smORF sequences exist in bacteria, yeast, plants and more complex organisms such as fruit flies, mice and humans (Ladoukakis *et al.* 2011). However, from these smORF sequences, the proportional rate of smORF peptide detection using traditional biochemical and proteomic techniques is very low. The question still remained as to how many smORF sequences are translated into functional proteins. Considering the presence of thousands of putative smORF genes (Flybase, Ensembl databases), the characterisation of three smORFs (*tal, Scl, Hemotin*) is unrepresentative of the entire pool that may exist overall. What is needed is the characterisation of a more representative sample of this new gene family. In order to conduct the functional characterisation of smORFs on a larger scale than before, we needed a comprehensive list of translated smORFs in which we have high confidence. This process is made difficult by the discovery of abundantly present non-coding RNA in the genomes of several organisms. These sequences are transcribed with RNA polymerase II and undergo post-transcriptional processing such as polyadenylation and splicing, just like normal protein coding transcripts, and many could contain putative smORFs, but there was previously no good method to make this differentiation.

**N.B. Portions of this chapter have been published in Aspden *et al.*, (2014) eLife.**
**For these experiments, J.L. Aspden designed the plasmids for tagged smORFs and cloning of plasmids was conducted by R.J. Phillips.**

Therefore, it is difficult to differentiate real ncRNA from peptide-encoding smORFs due to these shared characteristics, as the only thing that truly differentiates them is whether they are translated into protein. There are several evolutionary conserved genes that have been found to encode RNA rather than peptides. These non-coding RNAs are abundant in number and may be involved in a variety of cellular processes such as the processing of other RNA, transcriptional and translational regulation of genes and protein trafficking. An example of important mRNA-like ncRNAs includes *Pgc* RNA, which maintains germ cell fate in *Drosophila* by regulating transcriptional repression during early embryonic stages. *Drosophila* mutants of *bereft* ncRNA, usually expressed in the peripheral nervous system, exhibit malformed extrasensory organs and bristles found in the eye (Inagaki *et al.* 2005).

The results from Ribosome profiling experiments have recently led to the expansion of the putative proteomes of several organisms and the discovery of several novel aspects of ORF translation. The new technique of ribosome profiling has been conducted in yeast (Ingolia *et al.* 2009; Duncan and Mata 2014; Smith *et al.* 2014), *Drosophila* (Dunn *et al.* 2013), Mice (Ingolia *et al.* 2011) and zebrafish (Chew *et al.* 2013; Bazzini *et al.* 2014). In the case of identifying novel coding regions, ribosome profiling results indicate the presence of translated ORFs in previously unproductive genomic regions, including uORFs in the 5' UTR, as well as within 'non-coding RNAs' (ncr_ORFs) (reviewed by Ingolia 2014), as well as the discovery of alternative ORF translation (AltORFs) (Vanderperre *et al.* 2013). This method was employed by our lab to distinguish true coding smORFs from non-coding varieties of genes.

**Ribosome Profiling methods to determine translation**

The ribosome profiling method entails the Next Generation Sequencing of mRNAs protected by ribosomes. These are called 'ribosome bound fragments' (RBFs). RBFs are generated by nuclease digestion of mRNAs being transcribed in real time, thus allowing the identification of mRNA regions that are being actively translated by a ribosome. These short mRNA fragments (28-34nt) are then purified, sequenced and mapped to the transcriptome. This reveals the ribosome occupation along the mRNA, and shows actively translated regions of the transcripts, or where ribosomes may pile up and be

hindered from translating in some way. Sequenced mRNA fragments can then be quantified to also reveal the abundance of specific mRNAs that are undergoing translation (Ingolia *et al.* 2009).

The most recent releases of the *Drosophila* S2 cell genome (R5.50) upon writing of this thesis, showed the putative presence of 829 smORF peptide encoding genes, (~4% of total annotated protein-coding genes), which is double the predicted amounts in other metazoans. However, only 54 of these smORFs (6% of total pool) appear as candidates that we would consider for further study in the fly through some previous evidence of translation. These 54 smORFs each have 1) proteomics evidence, 2) a molecular GO term predicting a protein function from sequence homology and 3) coding sequence conservation in higher animals. These smORFs are often allocated a gene name based on protein homology of specific domains, but lack any specific functional data or in-depth study. A remaining 110 'partially corroborated smORFs' have two out of the three aforementioned hallmarks of translation. The remaining 665 annotated smORFs in *Drosophila melanogaster* have insufficient evidence of translation and most of these are identified by numerical cognate identifiers (CG numbers) rather than a gene name (Aspden *et al.* 2014). This amounts to a proportion of less than 20% (164/829) of 'Flybase-annotated' (FB) smORF genes with anything resembling solid evidence of translation in the fruit fly.

Given this lack of information, it is not surprising that *Drosophila* smORFs are considered untranslated, or potential lncRNAs at best. Better techniques were needed in order to assess smORF translation on a large scale. This endeavour led to the recently published work conducted by our lab, where a genome-wide assessment of smORF translation was performed, using an 'improvement' of the ribosome profiling method, which we call Poly-Ribo-Seq (Aspden *et al.* 2014). We addressed the question of non-productive ribosomal association by modifying the experimental protocol to add a Polysomal fractionation step prior to ribosome foot-printing (Arava *et al.* 2003). Polysomal fractionation allows separation of mRNA by the number of bound ribosomes, prior to nuclease digestion of the mRNA into RBFs which are subsequently purified and sequenced. By sequencing only the mRNAs bound to more than one ribosome, we hoped to improve the chances of detecting smORF mRNAs that

undergoing ‘active translation’ and in this case, active translation is inferred from the idea that more than one ribosome would not typically bind an mRNA in the absence of translation. Studies on yeast have shown that each translating ribosome spans a length of ~80nt (Arava 2003). It was therefore estimated that longer smORF mRNAs coding up to one hundred amino acids (303nt) can contain anywhere between 2-6 ribosomes. A potential caveat of sequencing mRNAs with two or more ribosomes is that it may instil a bias against dwarf smORFs such as *tarsal-less* (33nt) or *sarcolamban* (81nt), which would only need one ribosome to span the transcript. However, due to the observation that short smORF mRNAs seem to be enriched in smaller polysomes (<6 ribosomes), it could also be the case that ribosomes may be more densely packed in *Drosophila* (Aspden *et al.* 2014).

For an mRNA to be classified as translated in our Poly-Ribo-Seq experiments, stringent cut-offs were used. These were based on the subtraction of background signal from 3’UTRs of canonical protein-coding mRNAs (above the 90% percentile). The criteria for translation was 1) a minimum abundance of 11.8 RPKM (Reads Per Kilobase Per Million) to show that there is comparable abundance to canonical transcripts and 2) 0.57 Coverage (proportion of Poly-Ribo-Seq footprints covering the RNA transcript) because the higher the coverage, the more chances there are that the ORF is translated due to the binding of ribosomes across a longer stretch of the transcript (Aspden *et al.* 2014). Of the 274 smORFs transcribed in our S2 cells (from in-house RNA-seq of total S2 cell mRNA), 228 smORFs in total were found to be translated across two Poly-Ribo-Seq sequencing experiments.

The number of smORFs found to be translated by Poly-Ribo-Seq almost doubles the previous number of smORFs (164) with evidence of translation in *Drosophila* (proteomics evidence, a molecular GO term, coding sequence conservation) and gives a grand total of 285 smORFs with strong evidence of translation. There were also 2708 uORFs found to be translated in S2 cells, as well as 313 smORFs found in potential ncRNA transcripts. High confidence translated smORFs in ncRNA and 5’UTR regions (uORFs) would be interesting to follow up, as that would help to highlight whether there are any functional trends or differences between the smORFs found within these different transcripts.

**A requirement to corroborate translation of Poly-Ribo-Seq results**

Before ribosome profiling can become evidence of translation in lieu of a peptide product, previous ribosome profiling papers argue that there still exists a need to directly detect the peptide from novel protein-coding mRNAs found by ribosome profiling methods. The presence of non-AUG start codons, internal ribosome entry sites (IRES) and nonsense read through may hamper the correct prediction of peptide sequences (Ingolia *et al.* 2009). mRNA is subject to extensive post-transcriptional regulation, such as capping, splicing, polyadenylation and editing, which may or may not result in formation of a viable peptide product (Colgan and Manley 1997; Day and Tuite 1998; Li and Kiledgian 2010). mRNA can also be of varying stability, with a long or short half-life (Schoenberg and Maquat 2012). To prove that the high-confidence Poly-Ribo-Seq smORFs produce a peptide product, we had to independently demonstrate proof of translation. Independent monitoring of peptide translation would also show us whether our cut-offs for translation are meaningful. The Poly-Ribo-Seq experiment eliminates the question of active translation, but in order to provide evidence that this technique is indeed better than previous attempts, we decided to independently corroborate a representative pool of the 228 translated smORFs (at least 10%).

**Plasmid transfection and cellular imaging as a tool for smORF characterisation**

Engineered gene expression in *Drosophila* is a technique has been used previously by our lab to characterise smORF peptide function *in vivo* (GAL4-UAS system) and in cell lines using plasmid transfection (Galindo *et al.* 2007; Magny *et al.* 2013). The material used for Poly-Ribo-Seq by our laboratory was derived from *Drosophila* embryonic tissue culture (S2 cells) which are ideal for conducting sensitive RNA-based experiments as they are grown using aseptic methods. Using cells which can grow exponentially in a few days also provides abundant material for experiments of this type as there are several purification steps in the protocol, which leads to loss of some material as samples are carried over. Although there are almost 100 cell lines derived from *Drosophila melanogaster*, S2 cells are one of the most widely used cell lines for gene expression studies (Cherbas *et al.* 1994).

The *Actin5C* gene encodes cytoskeletal actin, which is present in all cell types. The proximal actin 5C promoter enables high levels of constitutive expression of downstream coding sequences (Chung *et al.* 1990). Plasmid transfection is a very powerful tool to introduce ectopic gene expression in *Drosophila* and other animal cell lines. The plasmid is combined with a lipid transfection agent to form membrane-bound liposomes, which are internalised by the cells. After internalisation, the plasmid is transcribed and translated by the host-cell machinery. (Lee *et al.* 2000; Wilder 2000). This technique allows us to ectopically express smORF peptides in S2 cells by transfecting the cells with a promoter and smORF-cDNA fusion plasmid, creating large amounts of expression of the target smORF. The main limitation of this approach is that the over-expression of some genes may lead to cell lethality, in which case we would fail to observe many cells expressing the construct (Wilder 2000).

This chapter entails the findings of the tagging-transfection assay designed to detect and corroborate Poly-Ribo-Seq smORF translation in S2 cells. This assay, coupled with the power of imaging, revealed specific patterns of smORF peptide subcellular distribution, which led us to assess the subcellular distributions and localisation of a larger sample of smORFs than have ever been tested before. These discoveries help us to shed some light on the potential functional roles of smORFs being expressed in *Drosophila* embryonic cells. This assay was conducted of smORFs across a range of different translation metrics, and amounted to 20% of the total smORFs deemed translated by Poly-Ribo-Seq.

# Results

## The tagging-transfection Assay

Verification of smORF translation *in vivo* requires contextual and developmental knowledge of when and where the mRNA expressed. Generating specific antibodies to each smORF would be a time-consuming and expensive endeavour that has not proven to be very successful in the past. In order to bypass this requirement, we decided to use inducible gene expression using plasmid transfection in S2 cells. By constructing a fusion plasmid containing a constitutively active promoter and smORF cDNA fused to a tag which can be labelled with a specific, efficient antibody, we can specifically label smORF peptides and image them at a cellular level.

Total S2 cell cDNA was used to clone individual smORF sequences from primers specific to each smORFs 5'UTR and coding sequence (CDS), minus the stop codon of the transcript. The full 5'UTR of the smORF was included to retain the Kozak sequence as well as any upstream translational regulation sequences. The cloned smORFs were inserted into a modified plasmid containing the actin 5C promoter, and a C-terminal tag (Figure 4.1A). The 3' end of the smORF lacking a 'stop' codon, is fused to the 5' of the tag so any expression of the peptide tag would be the result of translation of the smORF. This method can be repeated for a large number of smORFs, and can be used to provide corroboration of smORF translation in S2 cells.

## Experimental controls

We initially tested two different tags for the detection of smORF-fusion peptides. It was decided that we try the FLAG-tag, which is a small tag (amino acid sequence: DYKDDDDK) that can be readily detected by a very efficient commercially available antibody. The other tag we tested was Venus, which is a highly fluorescent tag that can be directly detected in cells without any subsequent labelling with antibodies. The use of a fluorescent tag could be useful in live-imaging, as we could potentially use it to observe translation or protein degradation in real time using live cells. Four constructs were initially designed as positive and negative controls for the experiment. Using site-directed mutagenesis, the ATG (methionine) at the start of the FLAG or Venus coding sequences was mutated to a GCG (alanine). These plasmids were called GCG-FLAG

(1) and GCG-Venus (2). These constructs served as the negative controls for the experiment, and should not be technically translated on their own as there is no start of translation included in these plasmids.

The positive controls for the experiment used the plasmids designed above, with the 5'UTR and first few codons of a translated smORF, CG42371 (Peptide Atlas Database) inserted upstream to introduce a non-plasmid AUG upstream of the GCG-FLAG and GCG-Venus sequences. These constructs were named AUG-mod-FLAG (3) and AUG-mod-Venus (4). If we observe translation of the tag from these constructs, it should be because of the non-plasmid start codon, such as those inserted from smORFs.

The AUG-mod-FLAG and AUG-mod-Venus constructs show FLAG and Venus signal in the cells, and this showed that the presence of a non-plasmid AUG leads to the translation of the tag, as expected. The GCG-FLAG plasmid showed no signal in the cells, however the GCG-Venus plasmid shows signal that is comparable to that present in the AUG-mod-Venus transfection, despite not containing a canonical start codon. As can be seen Figure 4.1B, the problem with the Venus fluorescent tag became apparent when these constructs were tested. We decided to clone the putative ORF within a known long non-coding RNA called Uhg2, into the GCG-tag constructs as an additional negative control, as Uhg2 is a snoRNA-host gene, which are a family of genes that have been shown to not be protein coding (Tycowski and Steitz 2001). Uhg2-mod-Venus showed signal in the cells, whilst Uhg2-mod-FLAG failed to produce a peptide that could be detected. Thus the reliability of the results from the Venus tagged constructs came further into question, as Venus was being translated in both of the negative controls.

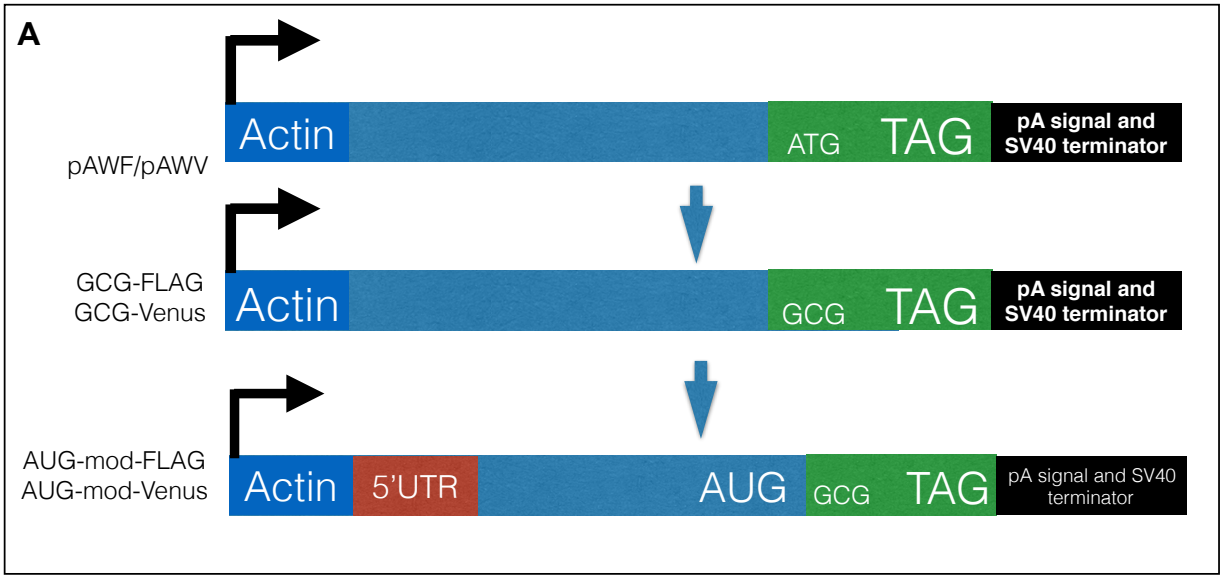
As each of the constructs had been sequenced to verify insertions and mutations, I looked into the GCG-Venus sequence and found that there are several downstream alternative start codons within the coding sequence of the Venus tag (GTG,Valine and CTG, Leucine). Western blots used to detect Venus and FLAG signal from these constructs showed the presence of multiple Venus protein bands corresponding to the size of the Venus tag alone, thus we concluded that there was translation of Venus from these alternative start codons (personal communication with MAS Mumtaz and JL Aspden). We therefore discontinued using the Venus tag in the tagging-transfection

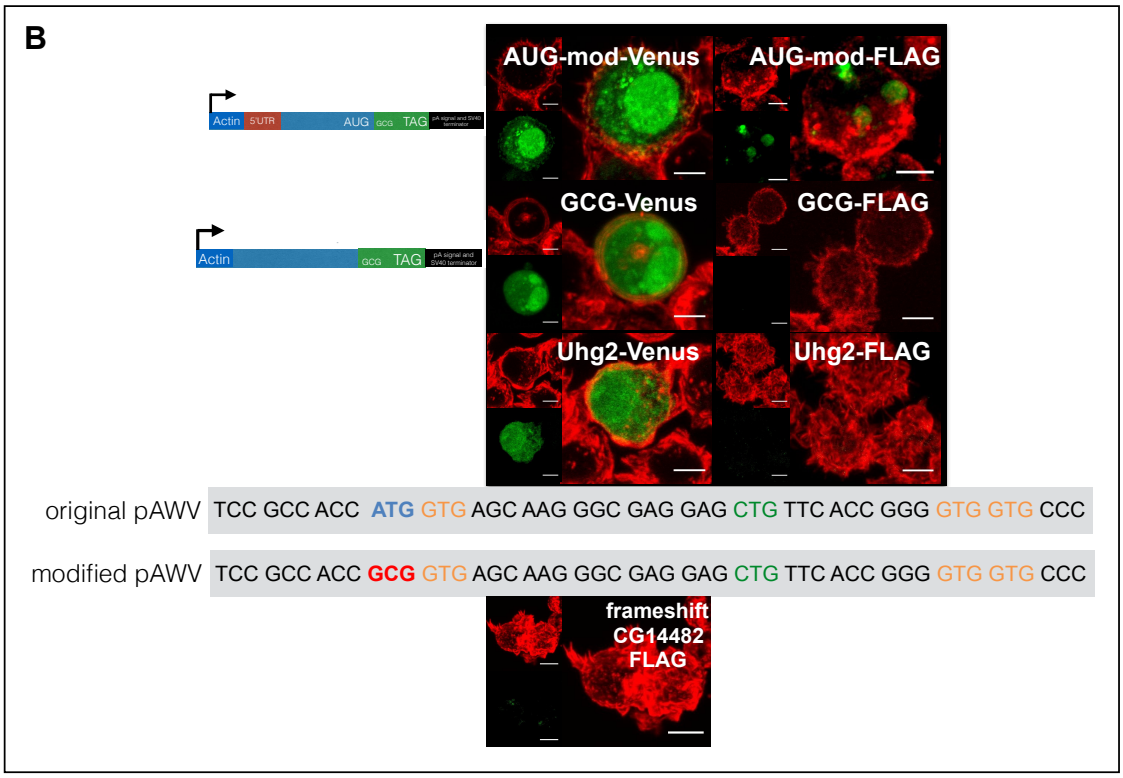
assay as it would potentially result in the quantification of false-positives. The remainder of the study was conducted using the FLAG tag as it was not translated on its own by any alternative start codons and also, using the mod-FLAG construct did not show translation of Uhg2, a known ncRNA. Transfections with a frameshift version of a translated smORF (CG14482) cds cloned into the GCG-FLAG plasmid did not result in any peptide being detected using antibody staining. This indicates that translation of the GCG-FLAG tag also does not occur when the coding sequence contains a nonsense mutation.

**Figure 4.1: Experimental controls for the tagging-transfection assay**

**A)** Modification of Murphy Lab Gateway Cloning expression vectors used in the transfection-tagging assay. Plasmid transcription is under control of the Actin5C promoter and the vector contains a C-terminal 3xFLAG/Venus tag sequence followed by an SV40 polyA signal. In order to eliminate expression of the tag only, the AUG start codon of the tag was changed to a GCG codon. These are termed GCG-FLAG and GCG Venus, and serve as a negative control in the experiment as there should not be translation of the tag in these constructs. The positive control AUG-FLAG/Venus construct was made by adding the 5'UTR plus the start codon of the translated smORF gene CG42371. These are termed AUG-mod-FLAG and AUG-mod-Venus.

**B)** S2 cells transfected with both negative and positive control constructs. In the case of FLAG constructs, cells were immune-stained to label FLAG (FITC secondary), followed by labelling of all samples with F-actin (Rhodamine-Phalloidin). Translation from both the positive control AUG-mod-FLAG/Venus constructs can be seen. No FLAG tag signal was observed for the negative control GCG-FLAG construct but GCG-Venus shows comparable levels of signal as the positive control. Testing the translation of a putative ORF in a known ncRNA, Uhg2 (Tycowski and Steitz 2001) reveals signal with the Venus tag, but not with the FLAG tag. Examining the first few in frame codons of the Venus tag reveals putative alternative start codons (orange) downstream of the original start codon (blue) and in the modified start codon (red). The presence of signal from the Venus tag in these negative controls rendered the Venus plasmids unusable to assess smORF translation in our assay. The bottom panel shows a frameshift construct of the translated smORF CG14482 to show that FLAG signal is specific to having an in-frame ORF. Scale bars = 5μm.


A
pAWF/pAWV
Actin
ATG
TAG
pA signal and SV40 terminator
GCG-FLAG
GCG-Venus
Actin
GCG
TAG
pA signal and SV40 terminator
AUG-mod-FLAG
AUG-mod-Venus
Actin
5'UTR
AUG
GCG
TAG
pA signal and SV40 terminator


B
AUG-mod-Venus
AUG-mod-FLAG
GCG-Venus
GCG-FLAG
Uhg2-Venus
Uhg2-FLAG
original pAWV TCC GCC ACC ATG GTG AGC AAG GGC GAG GAG CTG TTC ACC GGG GTG GTG CCC
modified pAWV TCC GCC ACC GCG GTG AGC AAG GGC GAG GAG CTG TTC ACC GGG GTG GTG CCC
frameshift
CG14482
FLAG

**smORF controls**

In order to assess the strength of the peptide tagging-transfection assay, we examined a smORF that is not normally translated, which is the last ORF of *tarsal-less* (*tal*). The *tal* transcript contains 5 ORFs (Galindo *et al.* 2007) and our lab have previously shown that ORFs 1A, 2A, 3A and AA are translated, whilst ORF-B is not translated, in a similar tagging-transfection experiment conducted in S2R+ cells (Figure 4.2A). A tal-ORFB mini-gene construct also does not rescue the tarsal-less mutant phenotype, indicating that this ORF is both nonfunctional and untranslated (Galindo *et al.* 2007) in the endogenous context of the gene. A new construct was designed for tal-ORFB for my study, and this plasmid contains the full-length *tal* transcript (ORF1A, 2A, 3A, AA and ORFB minus stop codon) inserted into the GCG-FLAG plasmid as described above. This construct did not appear to be translated, similar to the published observations of our lab and this result was verified several times (Figure 4.2B). Another *tal* construct was made, which essentially contained only ORFB, and lacks any of the upstream ORFs. This construct contains the inter-ORF sequence between the second-last and last ORFs of tal (ORFAA and ORFB), which serves as a 5'UTR for *tal-ORFB*. The sequence was then inserted in-frame with the GCG-FLAG tag. When tested, this tal-ORFB construct showed FLAG signal in the tagging-transfection assay (Figure 4.2C). The tal-ORFB transcript in the absence of upstream ORFs has not previously been tagged and tested in this manner. This result could mean that the presence of the upstream ORFs in dicistronic transcripts may 'soak up' the translation of downstream ORFs. *In vitro* translation has already shown that the amount of protein translated from each ORF decreases along the transcript according to the ORF position, with the 5'-most ORF1A showing the highest levels of protein (Galindo *et al.* 2007). The implications of this result for the corroboration of Poly-Ribo-Seq smORFs is that downstream dicistronic smORF ORFs should always be cloned into the GCG-FLAG plasmid in the context of their upstream ORFs to correctly determine whether they are translated.

As a negative control for the Poly-Ribo-Seq experiment, I tested the translation of four smORFs that had Poly-Ribo-Seq metrics below the cut-off for translated smORFs. CG33494, CecB (CG1878), CG13315 and CG34330 showed some ribosomal foot-

printing reads in one of the two Poly-Ribo-Seq experiments, but in the final analysis they fell below the high-confidence cut-offs for translation (Figure 4.2D). In the tagging-transfection assay, these smORFs showed FLAG signal comparable to that of smORFs that pass the translation filters. This result is also corroborated by Western Blotting, as protein could be detected from these transfections, but the signal was generally lower (MAS Mumtaz, personal communication). This result shows that smORFs below the cut-off for translation in the Poly-Ribo-Seq can be translated. This may mean that the filters for Poly-Ribo-Seq are too stringent or that these could be examples of false-positives in the tagging-transfection assay, resulting from the massive amount of induced expression with the Actin5C promoter, but it is clear that this promoter cannot 'force' the expression of something that lacks coding potential in the context of its endogenous gene model. It is more likely that the Poly-Ribo-Seq translation filters are the limiting factor, as other negative controls for translation, (frame-shift, ncRNA, tal-ORFB), do not show any signal in the cells. This makes the tagging-transfection a powerful tool as on a wider scale this result shows that if a translatable sequence is present, the tagging-transfection assay will detect the peptide with success.

**Testing the Poly-Ribo-Seq smORFs**

In the tagging-transfection assay, I tested several canonical smORF genes, dicistronic smORF genes, a number of long non-coding RNA ORFs as well as some uORFs that were deemed as translated in the Poly-Ribo-Seq experiments. At first, low-resolution light microscopy was used to assess whether FLAG signal, labelled by mouse anti-FLAG M2 antibody (SIGMA) followed by anti-mouse FITC secondary, could be observed in transfected cells (Outline of final method in Figure 4.3A). I was able to detect signal for all of the 19 FlyBase smORFs, 4 dicistronic smORF transcripts (8 smORF constructs), as well as the smORFs found in 6 lncRNA transcripts (12 constructs) and 6 smORFs found in 5'UTRs (uORFs) that were tested in the tagging-translation assay. The results of this experiment are summarised in Figure 4.3B. This provides proof of translation for the Poly-Ribo-Seq smORFs tested in this assay.

The result from tal-ORFB using a full length transcript versus just the ORFB sequence, showed us that true translation of downstream ORFs in dicistronic smORFs can only be reliably observed in the context of the upstream ORFs. Therefore, the ORF1 construct of the polycistronic smORFs contained just the 5'UTR but all downstream ORFs were tagged in a similar fashion to the full length tal-ORFB construct. Hence, two constructs were made for each of the dicistronic smORFs, i.e CG15386 is ORF1 to CG42371 and so on. Both ORFs in *Hemotin* (CG43210/CG43194) were found to be translated by our Poly-Ribo-Seq experiments, and are also corroborated by this assay. Interestingly, the polycistronic lncRNAs that were tested (pncr009:3L ORFs1-4, CR30055 ORFs 2,4-5), showed a similar result to that observed from the *in vitro* translation experiments of *tal* in Galindo *et al.* (2007); where decreasing levels of signal were observed for each subsequent downstream ORF (also see next section). uORFs (CG11284, CrebA, Sec22, FBtr0072084, FBtr0072210, FBtr0081720) and monocistronic lncRNA smORFs (CR32582, CR43144, CR43250) showed signal levels comparable to that of FlyBase smORFs. These results corroborate the translation of a significant proportion of high-confidence results of our Poly-Ribo-Seq experiments.

**Figure 4.2: Negative controls for the tagging-transfection assay**

**A)** The GFP tagging-transfection assay previously published by our lab. This experiment was conducted to determine *tarsal-less* ORF translation in *Drosophila* S2R+ cells. The results showed GFP signal from tagging of all ORFs except tal-ORFB, which is why it was chosen as a negative control for the new tagging-transfection assay. tal-ORFB has also been shown to not be functional (Galindo *et al.* 2007). **B)** FLAG-tagged expression of the full-length tal construct, tagging ORFB in S2 cells. *tal-ORFB* is not translated in the presence of upstream ORFs. **C)** When tal is cloned without the upstream ORFs, FLAG signal can be observed from the translation of the tal-ORFB sequence. **D)** Proof of translation of smORFs CG33494, CecB (CG1878), CG13315 and CG34330 in the S2 cell tagging-transfection assay, showing the sensitivity of this assay as it can detect the translation of smORFs that fall below the stringent translation cut-offs used in the Poly-Ribo-Seq data. Scale bars = 5μm.

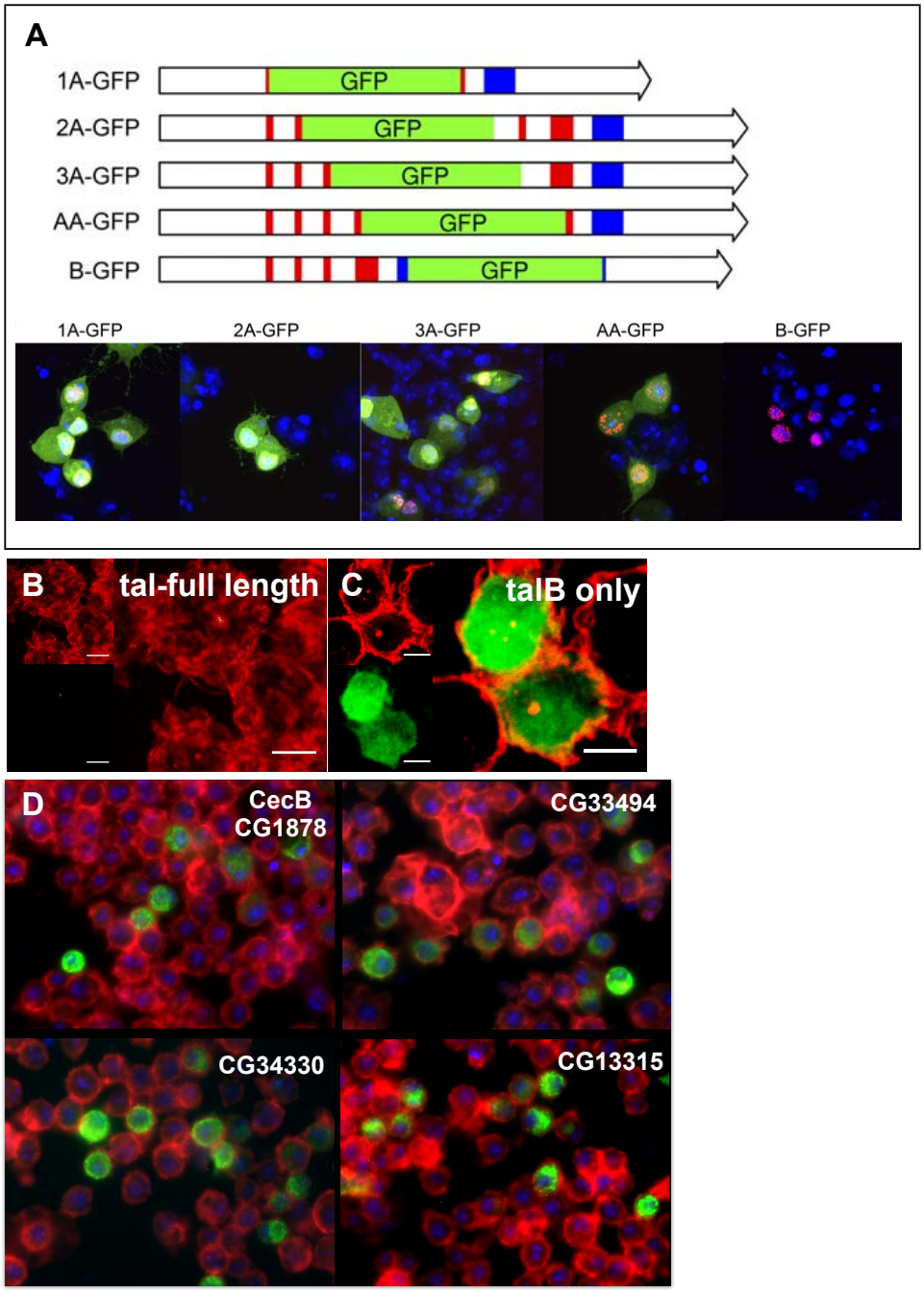
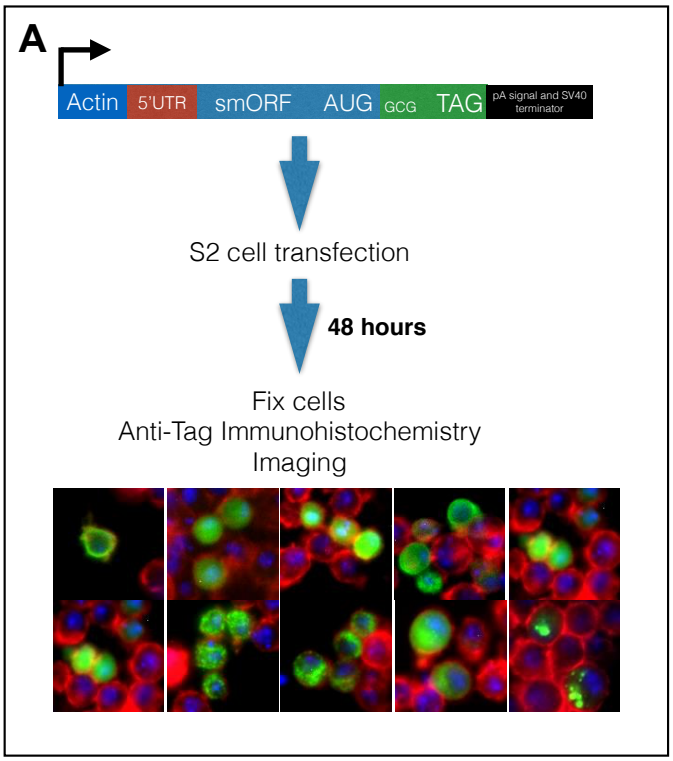

A
1A-GFP
2A-GFP
3A-GFP
AA-GFP
B-GFP
GFP
1A-GFP
2A-GFP
3A-GFP
AA-GFP
B-GFP
B
tal-full length
C
talB only
D
CecB
CG1878
CG33494
CG34330
CG13315

**Figure 4.3: Tagging-transfection assay results**

**A)** Overview of the tagging-transfection assay protocol: the 5'UTR plus the ORF sequence of the candidate smORF are cloned in-frame with the C-terminal tag in the GCG-FLAG plasmid and these are transfected in to S2 cells. These cells are harvested 48-hours post-transfection, fixed and an antibody stain using the FLAG M2 antibody, followed by anti-mouse FITC secondary, is performed before imaging. **B)** Summary of the tagging-transfection assay results. Tick marks indicate that signal was observed from FLAG-staining and Cross marks indicate that no signal was detected. This chart shows all 19 monocistronic smORFs, 4 pairs of dicistronic smORFs, 6 lncRNA smORFs (2 polycistronic and 1 monocistronic) and 6 upstream-ORFs tested are translated using this assay. The negative controls mod-FLAG, Uhg2-FLAG and full-length tal-B FLAG showed an absence of signal.




**B**

| smORF | | Polycistronic smORF | | lncRNA | | uORF | | Controls | |
|---|---|---|---|---|---|---|---|---|---|
| CG32230 | ✓ | CG15386 | ✓ | CR32582 | ✓ | CG11284 | ✓ | AUG-mod-FLAG | ✓ |
| CG44242 | ✓ | CG42371 | ✓ | CR43144 | ✓ | CrebA | ✓ | AUG-mod-Venus | ✓ |
| CG33170 | ✓ | CG42497 | ✓ | CR43250 | ✓ | Sec22 | ✓ | mod-FLAG | ✗ |
| CG12384 | ✓ | CG9878 Tim10 | ✓ | pncr009:3L (ORF1-4) | ✓ | FBtr0072084 | ✓ | mod-Venus | ✓ |
| CG32267 | ✓ | CG32736 | ✓ | CR30055 (ORF1,3,4) | ✓ | FBtr0072210 | ✓ | Uhg2-FLAG | ✗ |
| CG15456 | ✓ | CG43208 | ✓ | Rox2 (ORF1+ ORF2) | ✓ | FBtr0081720 | ✓ | Uhg2-Venus | ✓ |
| CG7630 | ✓ | CG43210 | ✓ | | | | | frameshiftCG14482 | ✗ |
| CG33199 | ✓ | CG43194 | ✓ | | | | | *tarsal-less* B full length | ✗ |
| CG33155 | ✓ | | | | | | | *tarsal-less* ORFB | ✓ |
| CG33774 | ✓ | | | | | | | Sarcolamban ORFA | ✓ |
| CG14482 | ✓ | | | | | | | | |
| CG34200 | ✓ | | | | | | | | |
| CG14036 | ✓ | | | | | | | | |
| CG6770 | ✓ | | | | | | | | |
| CG34242 | ✓ | | | | | | | | |
| CG33714 | ✓ | | | | | | | | |
| CG32276 | ✓ | | | | | | | | |
| CG14104 | ✓ | | | | | | | | |
| CG17278 | ✓ | | | | | | | | |
| | 19 | | 8 | | 12 | | 6 | | |

**Subcellular distribution of FLAG tagged smORFs**

The main advantage of using imaging to investigate smORF translation is that it presents a unique opportunity to observe other characteristics of the peptides besides translation. As I have previously demonstrated in flies using a GFP-tagged construct of Hemotin expressed in *Drosophila* haemocytes, this approach can allow us to assess potential function using the subcellular localisation of the tagged peptide. To observe the subcellular distribution of the FLAG-tagged smORF peptides in finer detail, transfected S2 cells were imaged using Laser Scanning Microscopy, which allows greater resolution and magnification of the samples. Using this imaging method, it became apparent that each smORF exhibited a specific arrangement within the cells. The smORFs were categorised into four classes of subcellular distribution based upon the pattern of tagged-peptide expression within the cells. Figures 4.4 to 4.9 show the labelling of smORF-FLAG (FITC, green channel) and F-actin (Rhodamine-Phalloidin, red channel) for the different smORFs that were tested in this experiment.

The most abundant class of tagged SEPs were the ‘Reticular’ SEPs (Figure 4.4), which are are distributed in a pattern of strings and puncta all over the cytoplasm, reminiscent of mitochondria. ‘Limited’ SEPs, such as those in Figure 4.5, are restricted to small and large punctae, sometimes surrounded by F-actin in the cytoplasm. ‘Other Cytoplasmic’ SEPs (Figure 4.6) appear as a diffuse signal all over the cytoplasm of the cells. FLAG-label showing some overlap with Rhodamine-Phalloidin were labeled as ‘Cortical’ SEPs (Figure 4.7).

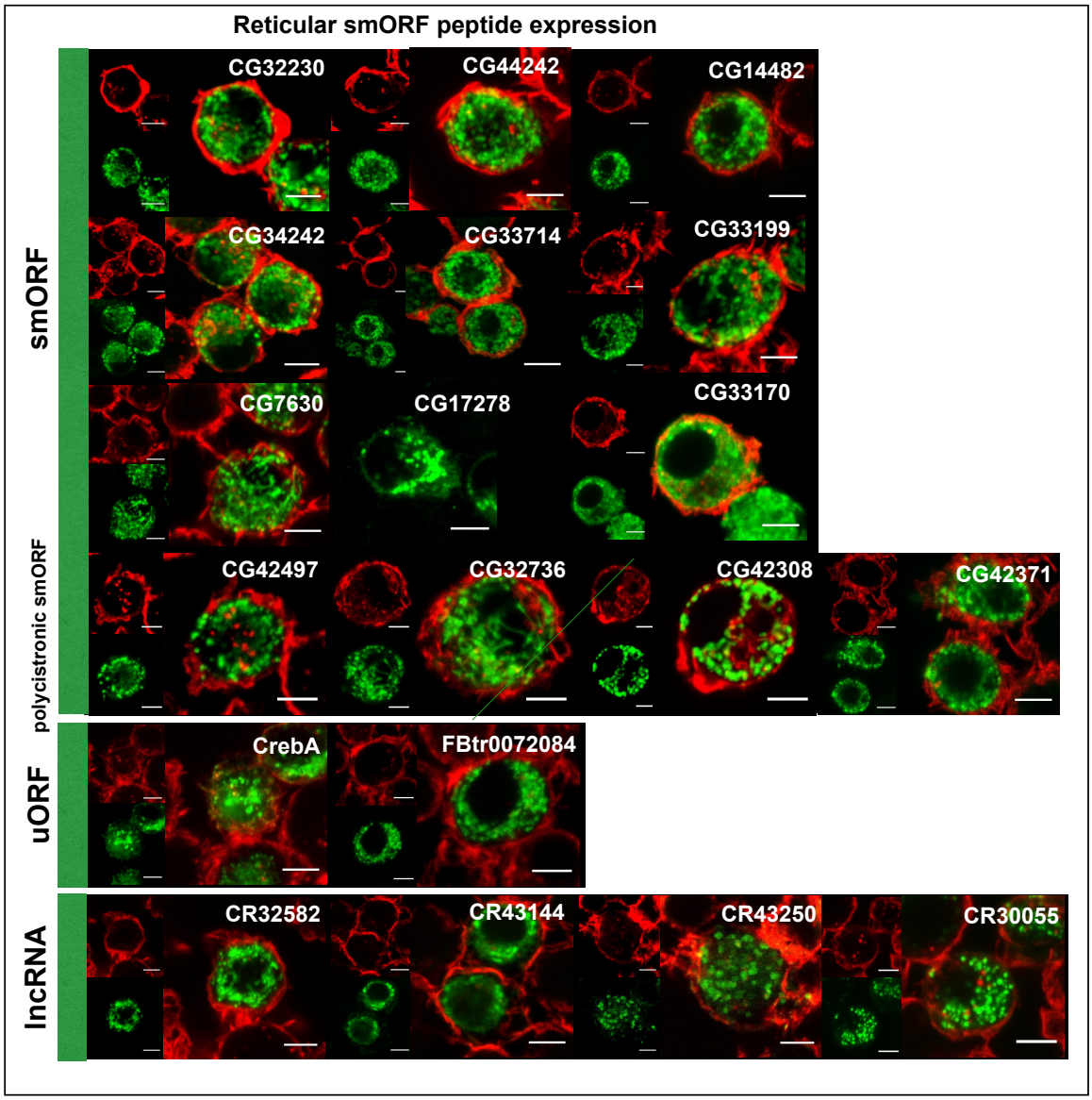



**Figure 4.4: FLAG-tagged SEPs with a 'Reticular' sub-cellular distribution**

The FLAG tagged peptide is labelled with FITC (green) while F-actin is highlighted by Rhodamine-Phalloidin dye. This most abundant class represents a mostly filamentous, reticulate distribution of peptides in a pattern similar to mitochondria. Scale bars = 5μm.

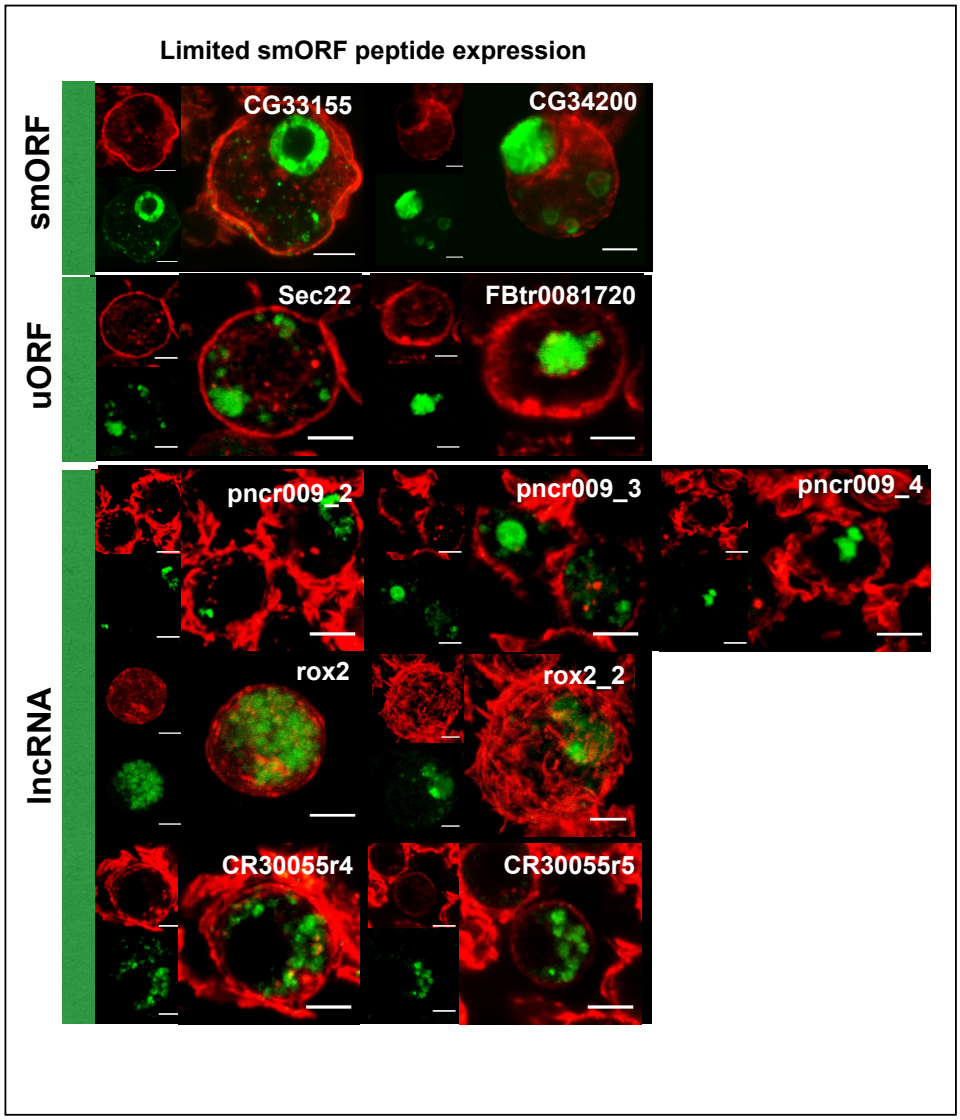



**Figure 4.5: FLAG-tagged SEPs with a 'Limited' sub-cellular distribution**

The FLAG tagged peptide is labelled with FITC (green) while F-actin is highlighted by Phalloidin dye conjugated to Rhodamine (red). This class represents SEPs with a a sub-cellular distribution that is limited to a small area of the cell and sometimes found in actin-enclosed spaces (CG34200, uORF FBtr0081720). Scale bars = 5µm.

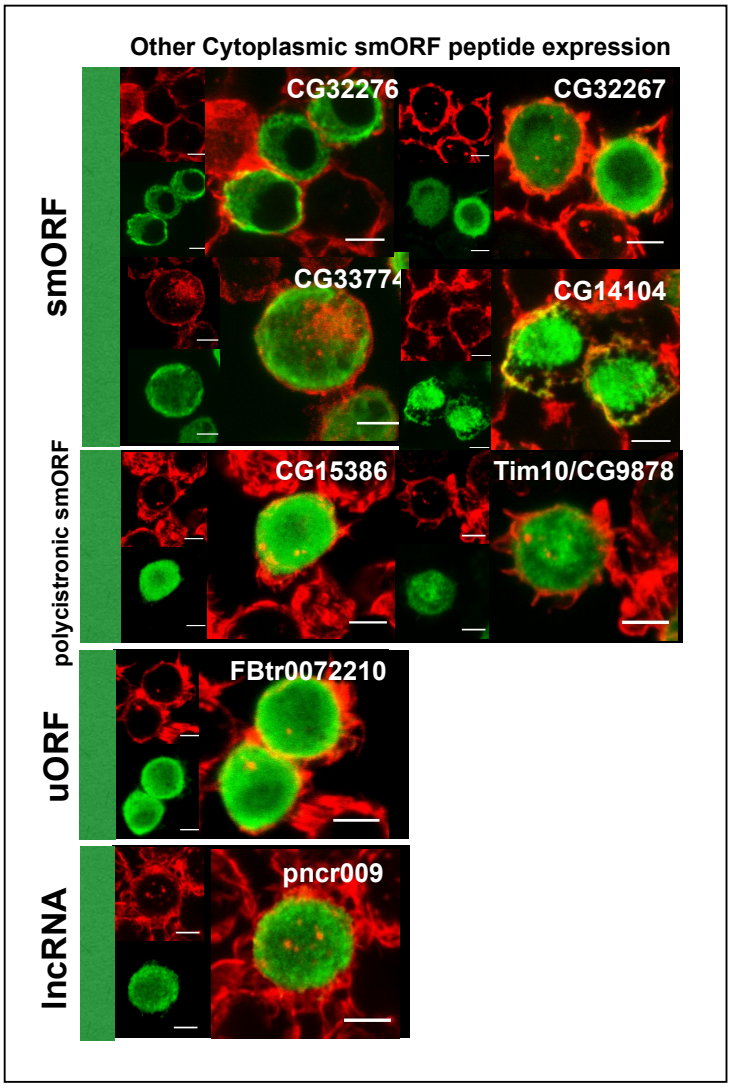



**Figure 4.6: FLAG-tagged SEPs with 'Other Cytoplasmic' sub-cellular distribution**

The FLAG tagged peptide is labelled with FITC (green) while F-actin is highlighted by Phalloidin dye conjugated to Rhodamine (red). This class represents SEPs with an even, non-specific distribution through the cytoplasm. Scale bars = 5μm.

Perhaps the Cortical SEPs, which include both of the Hemotin peptides, are involved in other aspects of endocytosis or intracellular trafficking. Overall, of the 45 smORF constructs tested, 19 are Reticular, 11 are Limited, 8 are 'Other Cytoplasmic' and 7 are Cortical. It was curious that the distribution of Hemotin ORFs were quite different with the FLAG-tag when compared to the expression found using the Hemotin-GFP construct in flies.

The high number of Reticular SEPs, as well as the distinct and varying patterns of the remaining classes, shows that the peptides have specific roles within the cell and are not just being translated due to a byproduct of the nature of the plasmid construct, which may have restricted expression of all of the SEPs in a similar manner. If smORF peptides are localising to specific organelles or other intracellular structures, this indicates that they may be functional, from a localisation point of view. The varying SEP localisation between the tested smORFs means the intracellular sorting of these peptides cannot be solely attributed to information found in the plasmid or the FLAG tag. Rather, SEP sequences contain their own information that help transport the peptide to their respective location in the cytoplasm (also see TargetP section below).

Cortical and Reticular SEPs are easy to distinguish, but the distributions of Limited and Other Cytoplasmic SEPs is harder to discern. It could be that the Limited SEPs are localised to an inclusion body or are being degraded in a degradation vesicle, which could be further verified by testing whether they are Lysosomal markers such as LysoTracker or the LAMP1 antibody, which labels the lysosomal membranes (Humphries *et al.* 2012). But the patterns of CG34200, CG33155, Sec22 and FBtr0081720 (uORFs) and pncr009 ORFs 2-4 (lncRNA) could also be associated with the nucleus since the peptides are found close to the centre of the cell. Due to the inability to conduct laser confocal imaging of the DAPI nuclear stain, I was unable to determine whether this is true. However, I believe that this experiment should be conducted in the future if resources and time permit.

**Overlap of SEP-FLAG with F-actin**

In order to see whether any of the SEPs tested show overlap with F-actin, Mander's colocalisation coefficient was used to quantify the overlap between SEP-FLAG

expression in transfected S2 cells, and the Rhodamine-Phalloidin labelled F-actin (Figure 4.7). Colocalisation analysis was performed for all smORFs having a minimum of 15 cells imaged in each transfection, with over three replicates. Overall, the co-localisation coefficients were quite low for overlap between SEPs and F-actin. However, the Cortical smORFs (CG12384, CG15456, CG6770, CG14036, CG43194, CG43210 and the uORF CG11284) have higher Mander's coefficients for overlap between FLAG peptide and Rhodamine-Phalloidin compared to the other subcellular distribution classes, but clearly the overlap does not account for all of the Cortical SEP that can be observed in the cells.

Two of the Cortical smORFs with considerable overlap to F-actin are the Hemotin peptides. As described earlier, Hemotin plays a role in phagocytosis, which depends on the F-actin cytoskeleton to conduct protrusion formation and endocytosis of material from the cellular environment. The colocalisation of Hemotin peptides to F-actin was previously not investigated in haemocytes, and it could be very likely that some of the peptide localises near actin, as per its proposed role in endosome maturation. As we did not have a positive control SEP to use as a positive control for localisation to F-actin, it is difficult to establish a cut-off in this case. However, if we take the lowest Mander's score of the Cortical smORFs as a cut-off (0.362), then all the other classes of distribution fall below any significant observable overlap to Phalloidin-Rhodamine. This shows that the visual observation of Cortical SEPs being localised to the cell periphery can be correlated to a numerical range of overlap of the smORF protein to a major cytoskeletal component.

**Overlap of SEP-FLAG with Mitochondria**

The most abundant class of Reticular SEPs exhibit a distribution pattern similar to mitochondria, therefore I decided to test whether these SEPs localise specifically to S2 cell mitochondria. This experiment required transfecting S2 cells with the smORF-FLAG constructs and labelling the mitochondria, which we chose to do with MitoTracker-RED dye (Life technologies). Cells were imaged using LSM at a magnification of 63x and Z-stacks obtained at an interval of 0.15-0.2um, which is the size range of of mitochondria (0.5-1um). Mander's colocalisation coefficient plugin was

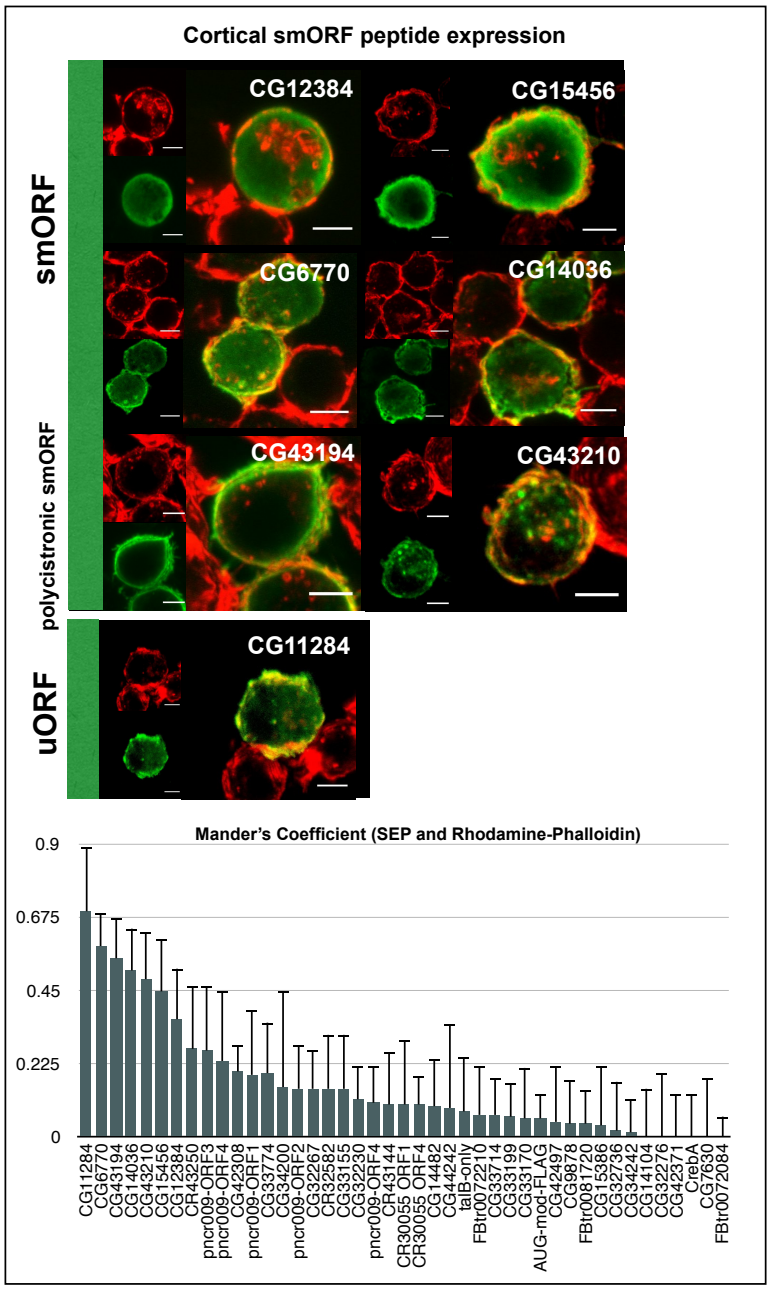



**Figure 4.7: FLAG-tagged SEPs with 'Cortical' sub-cellular distribution**

The FLAG tagged peptide is labelled with FITC (green) while F-actin is highlighted by Phalloidin dye conjugated to Rhodamine (red). Cortical SEPs tend to be localised to the periphery of the cytoplasm and thus have the highest amount of overlap with F-Actin, as can be seen from the graph. The y axis represents the 0-1 numerical value of overlap obtained by calculating averaged Mander's coefficients for 15 cells over three replicate transfections for each SEP. Scale bars = 5µm.

used to quantify the overlap between the SEP-FLAG expressed in the S2 cells, and the mitochondrial dye Mitotracker-RED (Figure 4.8). Colocalisation for Reticular SEPs, and some SEPs from other distribution categories for comparison, was quantified with a minimum of 15 cells imaged per transfection, in triplicate.

In order to establish a positive control for what amount of overlap can be considered to truly localise to MitoTracker-Red and thus mitochondria, I used haemocytes from larvae (L3) and white-prepupae (WPP) of a cross between flies containing *HemeseGAL4* and *UAS-mito-GFP (w[1118]; UAS-mito-HA-GFP.AP)2/CyO)*. This construct expresses a peptide with a mitochondrial-import signal fused to GFP. The haemocytes were fixed and stained using a similar protocol to that used with S2 cells and the samples were imaged as described above. The Mander's coefficients for mito-GFP and MitoTracker-Red signal in haemocytes at L3 (0.62±0.18) and WPP (0.74±.08) showed that mito-GFP and MitoTracker-Red did not colocalise completely, but still in quite high proportions. This control was chosen due to the availability of this construct in flies, and limited resources which prevented further cloning of plasmids. Otherwise, the ideal condition for the positive control would have been to transfect the mito-GFP construct as a plasmid in S2 cells under the control of the Actin5C promoter to ensure consistency between transfection conditions.

In order to establish a lower cut-off for mitochondrial colocalisation, as a negative control, we used the characterised SEP Sarcolamban (Scl) by tagging Scl-ORFA in the same GCG-FLAG plasmid as the tested smORFs. Scl-ORFA shows an 'other cytoplasmic' subcellular distribution, and had a medium (0.52±0.04) level of overlap to MitoTracker-Red. Scl peptide has been shown to be bound to and interacting with the Ca2+ ATPase pump SERCA in the sarcoplasmic reticulum (ER of cardiac muscles), as well as in the ER of *Drosophila* S2R+ cell line (Magny *et al.* 2013). The SERCA pump has also been shown to be present in the mitochondria of rat brown adipose tissue cells (De Meis *et al.* 2010), but we have never verified if this is the case in *Drosophila*. This made us consider that a partial overlap of Scl to MitoTracker-RED was not surprising as Scl would be found wherever the SERCA pump is required.

Alternatively, the amount of overlap between an ER protein and the mitochondria could be an artefact of imaging as standard laser confocal microscopy does not have the

resolving power to differentiate between structures closer than 0.1um. Hence, the relatively high score of Scl-ORFA to MitoTracker-Red may depict how closely tied the ER and Mitochondrial membranes are in S2 cells. Therefore, our cut-off for meaningful and total mitochondrial localisation of smORFs was chosen to be a minimum Mander's score of 0.52. Two Reticular uORFs (FBtr0072084 and CrebA) were not tested in this experiment, and there is one Reticular SEP (CG33170) that falls just below the 0.52 cut-off for Mitochondrial localisation. CG33170 could be partially associated with other inner-membrane networks that associate with the Mitochondria, such as the Mitochondrial Associated Membrane (MAM) of the Endoplasmic Reticulum (ER), or the ER itself. The Cortical and Limited SEPs that were tested had the lowest levels of colocalisation to MitoTracker-Red. This shows that by determining the subcellular distribution in these categories, I decreased the need for further testing of these SEP's for colocalisation to mitochondria.

**Analysis of Different Subcellular Distributions of SEPs**

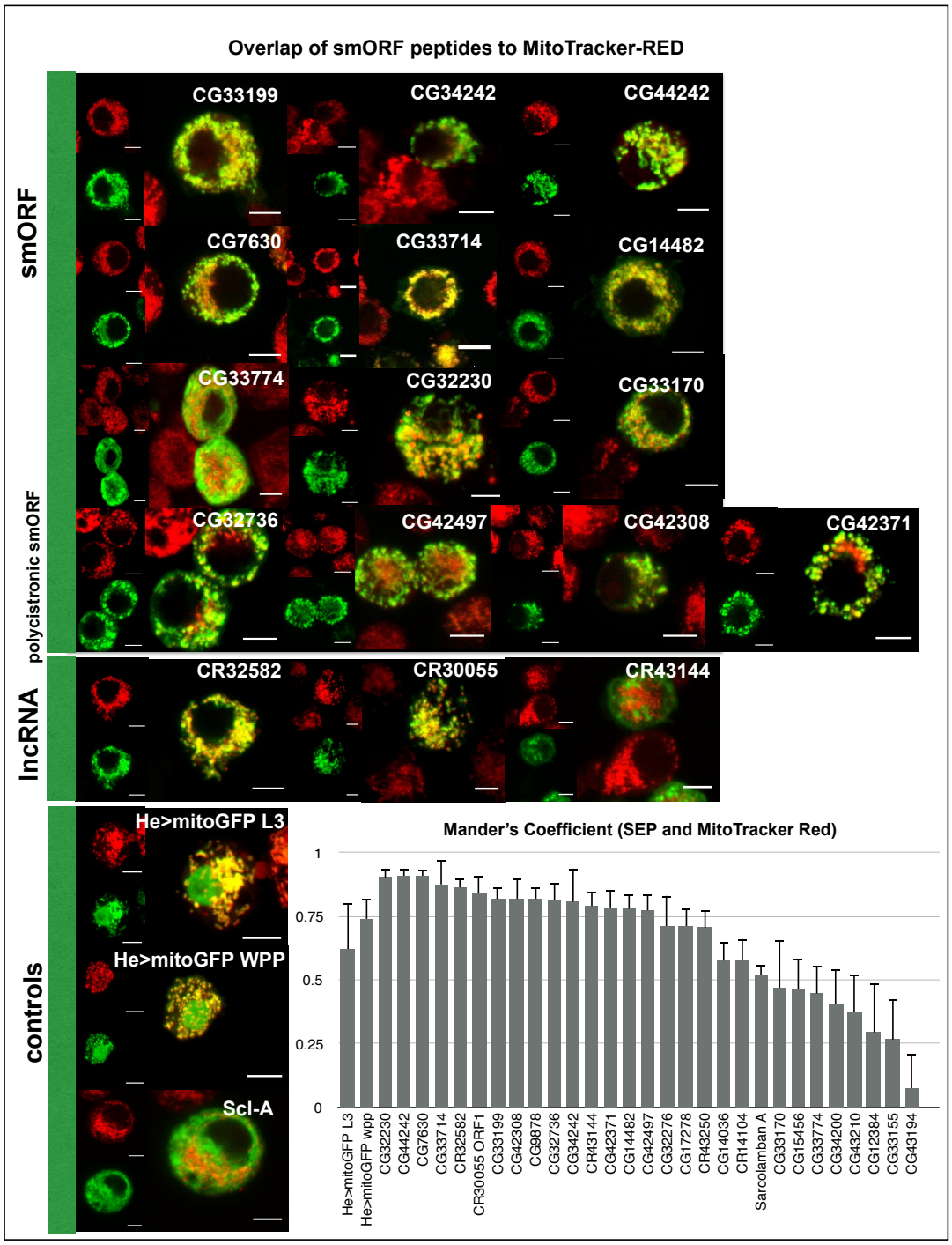
It should be noted that 64% of the Limited SEPs are encoded by smORFs present in long non-coding RNA transcripts. These include the downstream ORFs in pncr009:3L and CR30055. The SEPs encoded by the most 5' ORFs had a more widespread distribution of Other Cytoplasmic and Reticular, respectively, while the downstream ORFs of both of these lncRNA smORFs are classified as 'Limited'. The SEPs encoded by both ORFs of Rox2 are also Limited in their distribution. Generally, 50% of the SEPs tested in this analysis show a 'Reticular' sub cellular distribution (Figure 4.9A).

In order to test any correlation between SEP length and its sub cellular distribution, I compared the average peptide lengths of the different categories (Figure 4.9B). It is evident that Cortical and Reticular SEP's are larger in size, with median values of 73.71aa (± 29.14) and 66.79aa (±22.19) respectively, whilst Limited SEP's are shorter at35.38aa (± 18.84),and Other Cytoplasmic SEP's exhibit a wide range of sizes (53.00±27.39). Though these values may not be significantly different, this analysis suggests that the larger peptides tend to be Reticular or Cortical in distribution.

17 of the 18 Reticular SEPs tested show a high level of colocalisation to Mitotracker-Red, with average Mander's coefficients over 0.70, which means that these peptides

localise almost entirely to the Mitochondria. Determining subcellular distribution of SEPs using high resolution microscopy predicts their Mitochondrial localisation with high accuracy. However, there are three 'Other Cytoplasmic' SEPs (CG9878, CG32267 and CG14104) that localise quite highly to MitoTracker-Red, as well as one Cortical SEP (CG14036). These represent marginal cases like the previously mentioned CG33170, which could potentially fall either side of the cut-off. In total, this gives us 20 SEPs that show overlap to Mitotracker above the cut-off (Figure 4.9C). Altogether, categorising a SEP's subcellular distribution as Reticular confers a mitochondrial identity in most cases (Figure 4.9D), but the analysis is definitely improved and made more stringent when followed up by a localisation study and consistent imaging practice.

**Figure 4.8: MitoTracker-Red staining of S2 cells transfected with FLAG-tagged smORFs** Haemocytes from transgenic L3 and WPP animals expressing mitochondria localisation signal (*UAS-mito-GFP*) with haemocyte specific *HemeseGAL4* driver is used as a positive control for overlap with MitoTracker. In order to establish a cut-off value, S2 cells were transfected with Scl-ORFA, which has been been shown to interact with SERCA, a calcium pump that is present in ER (Mander's score 0.52). Using this method, the graph shows that out of the 28 SEPs tested, 20 have high overlap to MitoTracker-Red. Scale bars = 5μm.


Overlap of smORF peptides to MitoTracker-RED
smORF
polycistronic smORF
lncRNA
controls
CG33199
CG34242
CG44242
CG7630
CG33714
CG14482
CG33774
CG32230
CG33170
CG32736
CG42497
CG42308
CG42371
CR32582
CR30055
CR43144
He>mitoGFP L3
He>mitoGFP WPP
Scl-A
Mander's Coefficient (SEP and MitoTracker Red)
1
0.75
0.5
0.25
0
He>mitoGFP L3
He>mitoGFP wpp
CG32230
CG44242
CG7630
CG33714
CR32582
CR30055 ORF1
CG33199
CG42308
CG9878
CG32736
CG34242
CR43144
CG42371
CG14482
CG42497
CG32276
CG17278
CR43250
CG14036
CR14104
Sarcolamban A
CG33170
CG15456
CG33774
CG34200
CG43210
CG12384
CG33155
CG43194

The Poly-Ribo-Seq smORFs were also assessed for the presence of a transmembrane-alpha-helix (YC Eyre-Walker, unpublished data). Out of all the smORFs tagged and tested, 12 were predicted to contain a Hidden-Markov Model for a transmembrane alpha-helix (TMHMM) in the peptide sequence. Seven of the predicted TMHMM SEPs were found to localise to the mitochondria, showing an over-representation of mitochondrial SEPs in the overall TMHMM smORF sample. Conversely, 7 of 19 mitochondrial SEPs (37%) have predicted alpha helices (Figure 4.9E), meaning that these peptides may be embedded in the mitochondrial membrane. To confirm this, we would require higher resolution fluorescence microscopy, or more definitively, the much higher resolution technique of electron microscopy, to prove that the alpha-helical mitochondrial SEPs are actually spanning the mitochondrial membrane.

**TargetP analysis of subcellular localisation**

Translated peptides are sorted to various compartments depending on their N-terminal amino acid sequences, and correct subcellular localisation is essential for protein function (Schatz 1996). Signal peptides (SPs) motifs target proteins to the ER for processing through the secretory pathway, and have charged amino-acid residues close to a signal peptidase cleavage site in the amino acid sequence. Mitochondrial targeting peptides (mTPs) have a large proportion of Arginine, Alanine and Serine and usually lack negatively charged amino acid residues at the N-terminus (Schatz and Dobberstein 1996; Emanuelsson *et al.* 2000). TargetP is a program that was designed to predict whether a peptide sequence contains a SP or mTP motif and scores these at various confidence levels (Emanuelsson *et al.* 2000; Emanuelsson *et al.* 2007).

The TargetP localisation tool was used to analyse the peptide sequences of different smORF data sets in order to predict the localisation of the peptide from smORF N-terminal sequences (Figure 4.10). 805 of the 829 ‘annotated’ FlyBase smORFs were run through the TargetP program and only 79 (9.8%) of these are predicted to localise to Mitochondria. However, 351(43.5%) Flybase smORFs are predicted to contain a Signal peptide (SP) motif, which is quite a large proportion of all *Drosophila* smORFs. Of the 228 smORFs peptides found to be translated by Poly-Ribo-Seq in S2 cells, 39 (17.5%) are predicted to be Mitochondrial, and 43 (19.3%) are predicted to be Signal peptides.

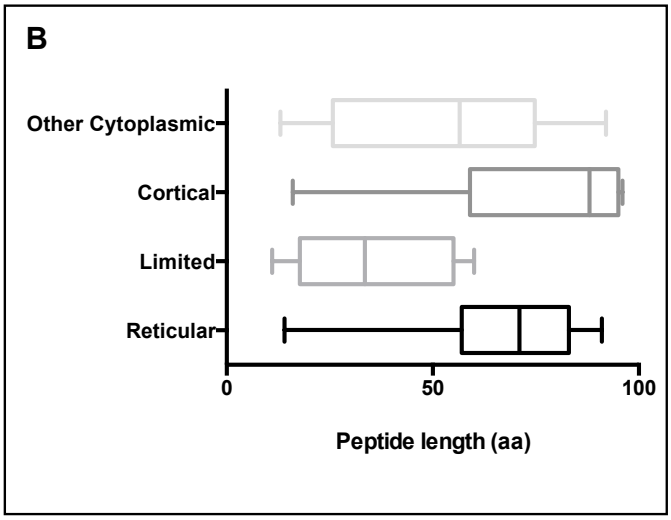
TargetP analysis of the subset of Poly-Ribo-Seq smORFs that were tested in the tagging assay shows that 22% contain mTP motifs and 15.6% contain SP motifs. A shortcoming of TargetP is erroneous high-confidence prediction, as can be seen from our tagged pool. Of the 20 mitochondrial SEPs, only half were predicted to have mTPs by TargetP, of the remaining SEPs, 3 were predicted to have SP motifs, which is incorrect. This shows that bioinformatics predictions alone are not a reliable method to asses the subcellular localisation of SEPs. However based on my own experimental data showing overlap between Reticular smORFs and MitoTracker-Red; I have shown that the proportion of mitochondrial smORFs is actually closer to 44% of the entire smORF sample tested. If we extrapolate the data from the tested smORFs to all Poly-Ribo-Seq smORFs, it could mean that up to half of the Poly-Ribo-Seq smORFs are mitochondrial, which would be very surprising, but it was also very surprising that such a high proportion of our tested sample are mitochondrially localised. Considering that S2 cells are an immortalised cell line, it would not be a far flung assumption to conclude that homeostatic and cellular function related smORFs are the more abundant ones, but in order to conclude if this is a representative pool of the entire sample, we would need to tag even more of the translated S2 cell Poly-Ribo-Seq smORFs. The complete results of all the smORFs tested in the tagging-transfection assay are summarised in Figure 4.11.

**Figure 4.9: Summary of tagged-SEP subcellular distributions**

**A)** Distribution of the subcellular localisation of each category of smORFs. Overall, Reticular SEPs are the most abundant class of smORFs (42%) followed by Limited SEPs (24%) **B)** Box plots showing the median peptide length for each sub-cellular distribution class of smORFs. The Cortical (median-88 aa) and Reticular (median-71 aa) SEPs tend to have a longer peptide length, while the Limited class has the shortest median length (34 aa), possibly due to the abundance of shorter ncr_ORF 'dwarf smORFs'localisation in this class. **C)** List of smORF genes whose SEP's show mitochondrial localisation; the majority (89%) of these exhibits a Reticular distribution D) Overlap between the mitochondrial SEP's and those SEP's predicted to contain an alpha helix by the prediction program TMHMM, there is no significant correlation and only 37% of the mitochondrial SEP's are predicted to contain a transmembrane alpha-helix.

| A | smORF | polycistronic | lncRNA | uORF | | |
|---|---|---|---|---|---|---|
| Reticular | 9 (48%) | 4 (21%) | 4 (21%) | 2 (10%) | 19 | 42.2% |
| Limited | 2 (18%) | 0 | 7 (64%) | 2 (18%) | 11 | 24.4% |
| Other Cytoplasmic | 3 (37%) | 2 (25%) | 1 (13%) | 2 (25%) | 8 | 17.8% |
| Cortical | 4 (57%) | 2 (29%) | 0 | 1 (14%) | 7 | 15.6% |

B

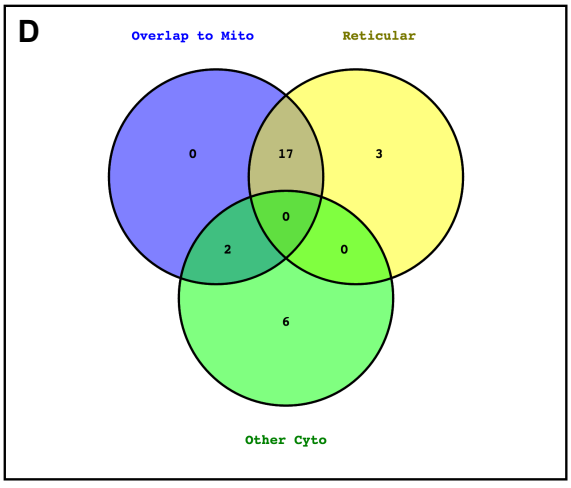



C

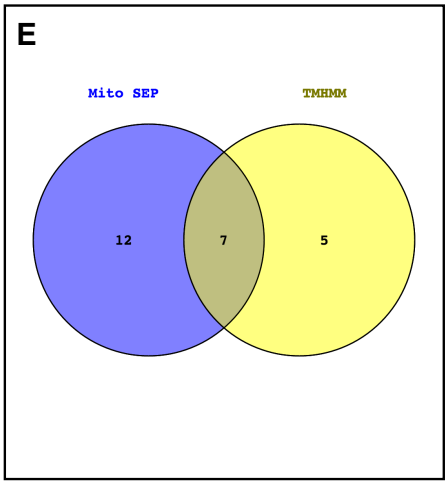
**Mitochondrial SEPs**

| | |
|---|---|
| CG32230 | CG14482 |
| CG44242 | CG42497 |
| CG7630 | CG32276 |
| CG33714 | CG43208 |
| CG33199 | CG17278 |
| CG34242 | CR14104 |
| CG42308 | CR43144 |
| CG9878 | CR14036 |
| CG32736 | CR32582 |
| CG42371 | CR30055 (ORF1) |

D




E



| | FB smORF | Poly Ribo Seq smORF | Tagged smORF |
|---|---|---|---|
| mTP | 79 (9.8%) | 39 (17.5%) | 10 (22.2%) |
| SP | 351 (43.5%) | 43 (19.3%) | 7 (15.6%) |
| Not determined | 376 (46.7%) | 141 (63.2%) | 28 (62.2%) |

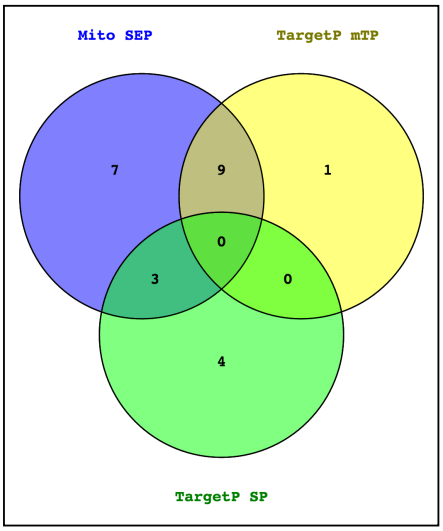



**Figure 4.10: TargetP predictions for all translated smORF categories**

smORF peptide sequences were run through the TargetP program to predict for the presence of a putative signal peptide (SP) or mitochondrial localisation (mTP) motifs in the complete pool of annotated smORFs (FB smORFs), the Poly-Ribo-Seq candidate smORFs and the pool of tagged smORFs. The MitoTracker colocalisation results show the localisation of an additional 7 SEP's that are not predicted to have either a SP or mTP motif.

**Figure 4.11: Summary table of all the tagged smORFs**

This table shows the data previously known about these smORFs, such as SEP length (aa), TMHMM prediction, Poly-Ribo-Seq translation and prior peptidomic evidence. There is now additional data provided by the tagging-transfection assay, such as translation in the transfection-tagging assay, subcellular distribution class and results from the co-localisation analysis with Phalloidin-Rhodamine (F-Actin) and MitoTracker Red along with the TargetP predictions of SP and mTP motifs.

| smORF | Length (aa) | Tm alpha helix | Poly Ribo Seq Tranlated | Peptidomic Evidence | FLAG-tag translation | TargetP | Distribution | Phalloidin-Rhodamine | MitoTracker-Red |
|---|---|---|---|---|---|---|---|---|---|
| CG32230 | 83 | 1 | Yes | Pep. Atlas | Yes | - | Reticular | Low | High |
| CG44242 | 70 | 0 | Yes | Couso Lab | Yes | Mito | Reticular | Low | High |
| CG33170 | 71 | 0 | Yes | No | Yes | Signal | Reticular | Low | Medium |
| CG32267 | 49 | 1 | Yes | Pep. Atlas | Yes | Mito | Other Cytoplasmic | Low | Not tested |
| CG12384 | 96 | 0 | Yes | Pep. Atlas | Yes | - | Cortical | Medium | Low |
| CG15456 | 95 | 0 | Yes | Pep. Atlas | Yes | - | Cortical | Medium | Medium |
| CG7630 | 90 | 1 | Yes | Pep. Atlas | Yes | Mito | Reticular | Low | High |
| CG33199 | 79 | 0 | Yes | Pep. Atlas | Yes | Mito | Reticular | Low | High |
| CG33155 | 60 | 1 | Yes | Pep. Atlas | Yes | - | Limited | Low | Low |
| CG33774 | 40 | 1 | Yes | Pep. Atlas | Yes | Signal | Other Cytoplasmic | Low | Medium |
| CG14482 | 57 | 0 | Yes | Pep. Atlas | Yes | Mito | Reticular | Low | High |
| CG34200 | 52 | 0 | Yes | Pep. Atlas | Yes | - | Limited | Low | Medium |
| CG6770 | 69 | 0 | Yes | Pep. Atlas | Yes | - | Cortical | Medium | Not tested |
| CG34242 | 58 | 1 | Yes | No | Yes | Signal | Reticular | Low | High |
| CG33714 | 90 | 0 | Yes | Pep. Atlas | Yes | Mito | Reticular | Low | High |
| CG32276 | 64 | 1 | Yes | No | Yes | Mito | Other Cytoplasmic | Low | High |
| CG14104 | 68 | 0 | Yes | No | Yes | - | Other Cytoplasmic | Low | High |
| CG14036 | 93 | 0 | Yes | Pep. Atlas | Yes | - | Cortical | Medium | High |
| CG17278 | 80 | 0 | Yes | No | Yes | - | Reticular | Low | High |
| CR43144 | 91 | 0 | Yes | No | Yes | - | Reticular | Low | High |
| CR32582 | 52 | 0 | Yes | No | Yes | Mito | Reticular | Low | High |
| CR43250 | 73 | 0 | Yes | No | Yes | - | Reticular | Low | High |
| pncr009:3L | 1-21, 2-20 3 - 33 | 0 | Yes | No | Yes | - | Other Cytoplasmic + Limited | Low | Not tested |
| Rox2 | 1 - 38, 2-17 | 0 | Yes | No | Yes | - | Limited | - | Not tested |
| CR30055 | 2-53, 4-17, 5-56 | 0 | Yes | No | Yes | - | Reticular + Limited | Low (ORF1) | High (ORF1) |
| CG11284 | 16 | NA | Yes | No | Yes | - | Cortical | Medium | Not tested |
| CrebA | 16 | NA | Yes | No | Yes | - | Reticular | - | Not tested |
| FBtr0072084 | 14 | NA | Yes | No | Yes | - | Reticular | Low | Not tested |
| FBtr0072210 | 13 | NA | Yes | No | Yes | - | Other Cytoplasmic | Low | Not tested |
| FBtr0081720 | 11 | NA | Yes | No | Yes | - | Limited | Low | Not tested |
| CG15386 | 77 | 0 | Yes | No | Yes | Signal | Other Cytoplasmic | Low | Not tested |
| CG42371 | 87 | 0 | Yes | No | Yes | Mito | Reticular | Low | High |
| CG42497 | 65 | 0 | Yes | No | Yes | - | Reticular | Low | High |
| CG9878 | 92 | 1 | Yes | No | Yes | Signal | Other Cytoplasmic | Low | High |
| CG32736 | 79 | 1 | Yes | No | Yes | Mito | Reticular | Low | High |
| CG42308 | 61 | 1 | Yes | No | Yes | Signal | Reticular | Low | High |
| CG43194 | 88 | 1 | Yes | Pep. Atlas | Yes | Signal | Cortical | Medium | Low |
| CG43210 | 59 | 1 | Yes | Pep. Atlas | Yes | Mito | Cortical | Medium | Medium |
| tal-B | 49 | | Yes | No | Yes | - | Other Cytoplasmic | Low | Not tested |
| tal-B full lengt | 49 | | No | No | No | - | N/A | not tested | Not tested |
| Scl-A | 28 | | No | Yes | Yes | - | Other Cytoplasmic | not tested | Medium |

# Discussion

Polysomal Ribosome profiling is a new method being used to describe smORF translation, therefore we decided to independently corroborate translation of Poly-Ribo-Seq smORFs using heterologous gene expression in S2 cells. This would require the detection of the peptide product that is being encoded by a smORF. It is difficult to make a good specific antibody against a smORF peptide, as we have found in the case of Tarsal-less, Sarcolamban and Hemotin. This presents a challenge in detecting the endogenous gene product in wild type animals, therefore we chose to circumvent the problem by conducting a tagging-transfection assay in S2 cells. Thus we employed the use of a cleverly designed smORF-tag-fusion plasmid to help independently corroborate that smORF genes are able to produce a peptide product.

**Independent corroboration of translated smORFs**

Two different tags were used to help label the SEPs, these are: Venus, a bright, modified YFP tag, and FLAG which is a synthetic peptide readily detected by an efficient antibody (Einhauer and Jungbauer 2001). Testing negative controls with both tags revealed that Venus signal is detected in cells which in theory, should not translate the tag due to the mutation of the start codon from ATG to GCG. It was postulated that the abundance of the transfected plasmid, due to the Actin promoter, leads to the translation of the Venus peptide from alternative start codons. This has has been shown in the case of NAT1 which uses GUG as a translation initiation codon in *Drosophila* (Takahashi *et al.* 2005). Rather than also mutating all further downstream start codons in the modified Venus plasmid, the Tagging-transfection Assay was continued with the constructs containing modified-FLAG tag, which contains no putative downstream start-codons in its sequence. The strengths of the tagging-transfection assay were highlighted by the results from the testing of tal-ORFB using different cloning conditions. As tal-ORFB did not rescue tal mutant phenotypes, and is not as conserved as the other tal ORFs, it was previously deemed to be untranslated and nonfunctional (Galindo *et al.* 2007). My results corroborate these findings, as when tal-ORFB is cloned in its *in vivo* context, with all the upstream ORFs present in the transcript. I did not observe translation of this ORF, despite the massive over-expression of the mRNA that would occur due to the

actin 5c promoter. When the upstream ORFs are removed in the plasmid, FLAG signal can be seen from the translation of tal-ORFB. This shows that in the tagging translation assay, when entire smORF transcripts are cloned in the context of their endogenous 5'UTR and upstream coding regions, they will only be translated if they have translational potential.

For all smORFs that appear to be translated using Poly-Ribo-Seq, as well as some that fall below the translation cut-off, we could visualise the FLAG-tagged SEP using fluorescent antibody staining of S2 cells. The fact that smORFs below the translation cut-offs also appear translated does not necessarily mean that they are false-positives since closer inspection of the Poly-Ribo-Seq data reveals that these smORFs were found to be translated in one repeat of the Poly-Ribo-Seq experiments, but their RPKM fell below the cut-off when averaged over all the repeats. Although the smORFs are probably being massively over-expressed by the Actin5C promoter, the tagged smORF constructs are derived from the cDNA and gene models from the same *Drosophila* cell line used for the Poly-Ribo-Seq experiment. The detection of a peptide product from these smORF sequences, and evidence of their translation from Poly-Ribo-Seq, shows that this sub-population of smORFs are definitely translated by two independent methods. The real lncRNA Uhg2 is not expressed under these same conditions, despite the constitutively active nature of the Actin5C promoter and being under the context of its own 5'UTR. Additionally, Western Blotting of S2 cells transfected with the negative control smORF constructs failed to detect peptide products by that method (MAS Mumtaz, personal communication; Aspden *et al.* 2014).

The process of detection through imaging has emerged as more sensitive in the detection of SEPs as it allows the researcher to search by eye for the cells that contain fluorescent signal, as opposed to using western blotting, which would need a large proportion of cells expressing the construct to counteract any possible effects of low transfection efficiency as well as loss of material over the several steps involved in the immunoblotting protocol. By extrapolation, the 228 high-stringency Poly-Ribo-Seq smORF hits can therefore be believed to be translated, due to the work presented here, where I have corroborated the translation of a representative selection of almost 12% of the entire translated pool.

Most of these high-stringency smORFs have no previous proof of function or translation and are therefore ideal novel candidates to screen for further characterisation. These results are significant in our claim that smORFs are a special class of small peptide-encoding genes and that there exists huge potential for several thousand other smORF sequences in the genome to also be translated. The fact that smORFs which fall just under the translation metric cut-offs are also translated, shows that our metrics for assessing translation for Poly-Ribo-Seq are probably too stringent. However, some smORFs may just be transcribed at lower levels in normal, wild type S2 cells. It may be that we have not have sampled the required cellular context or experimental condition allowing for the translation of the many more smORF transcripts that exist in the genome. The Poly-Ribo-Seq experiment should therefore be repeated under different cellular growth conditions, similar to the modENCODE RNA-Seq data, to include conditions like infection, toxification, as well as nutrient stress or starvation. This would reveal smORFs which may be involved in the cellular stress response, as opposed to those involved in normal cellular homeostasis which may be the case for these high stringency smORFs translated in wild type S2 cells.

**Functional assessment of SEPs by Subcellular Distribution and Colocalisation**

Imaging of FLAG-tagged SEPs not only allowed more sensitive detection of these peptides than immunoblotting but also enabled us to assess characteristics of their potential cellular function. The transfected smORFs showed distinct subcellular distributions in the cytoplasm of the S2 cells, indicating that they may be performing specific functions. Though using a plasmid with a strong constitutive promoter may have exaggerated levels of transcription of these smORFs, the display of four distinct types of subcellular distributions of the translated SEPs indicates that the higher levels of transcription did not affect the expression of smORFs in such an aberrant manner. as even if the cell is overwhelmed by the smORF transcript, it seems to maintain the homeostasis which allowed the peptides to be shuttled to various subcellular compartments, and the peptide can be found in enough cells to be seen and imaged. The presence of the FLAG tag also does not seem to make the SEP behave in any one particular fashion, and despite the presence of this additional peptide, the cellular

machinery is able to conduct the post-translational modifications needed to sort the individual peptides to their distinct locations. If the question arises of whether translation of these peptides resulted in cell stress or abnormality, this can be negated by the fact that the peptides were readily detected in the transfected samples over several repeats, showing that smORF encoded peptides are stable and non-toxic in the cells.

This encouraging result led us to classify the SEPs into subcategories of localisation and distribution in order to postulate their potential functions. Some FLAG tagged SEPs showed overlap with Rhodamine-Phalloidin marker which was used to label the cell periphery, these were labelled as ‘Cortical’ SEPs. There are similar numbers of ‘Other Cytoplasmic’ SEPs, which show a hazy network of peptide distribution all over the cell body. The ‘Limited’ SEPs were highly represented by ncRNA smORF types, and their restricted localisation may lend an explanation as to why they are detected at lower levels in Poly-Ribo-Seq experiments. It was revealed by imaging that they tend to be localised in a single or just a few small punctae, bound by F-actin in some cases, and found close to the centre of the cell. It is generally accepted that mRNA transcripts are shuttled close to regions in the cell where they are required (Brar *et al.* 2015). It may be that smORFs located in putative ncRNAs are generally lowly translated, or perhaps they are sampled at lower levels because they are associated to the nuclear region (which is excluded from RNA-profiling by crude fractionation) (Aspden 2014).. In either case, this hypothesis should be tested as it would provide some insight as to whether the spatial domain of a smORF transcript may result in variations in detection and also confirm whether Limited SEPs are localised to the nucleus. This could be done by checkcing overlap of the peptide with nuclear stain such as Hoechst.

By labelling the cellular F-actin with Rhodamine-Phalloidin dye, SEP overlap to this major cytoskeletal component could be quantified using Colocalisation analysis (Mander’s overlap coefficient). This revealed very low levels of overlap values of Rhodamine-Phalloidin with most SEPs, but for all the SEPs that were in the ‘Cortical’ category, the coefficient is above 0.3. This indicates that not all of the ‘Cortical’ SEP localises to F-actin, but a partial spatial overlap can definitely be observed specifically for these peptides in relation to the colocalisation coefficients of the‘Reticular’, ‘Other Cytoplasmic’ and ‘Limited’ SEPs. Hemotin was also tested in the tagging-transfection

assay, and as expected, both of the ORFs produced a tagged peptide product. Interestingly, the 88aa ORF1 was found mainly at the cell edge, and the 59aa ORF2 localised to the cell edge as well as in puncta around the cytoplasm. Therefore, it would be interesting to see whether the other 'Cortical' smORFs are also involved in endocytosis or immunity, based on their similar localisation to Hemotin.

**Discovery of novel mitochondrial SEPs**

The most abundant category of smORFs were the 'Reticular' smORFs, which displayed a pattern reminiscent to mitochondria. The punctate and stringy distribution of the Reticular smORFs led to the hypothesis that the Reticular smORFs are localising to mitochondrial structures in the cell. This hypothesis was tested by transfecting S2 cells with the smORF-FLAG constructs and labelling mitochondria with MitoTracker-Red dye. The signal overlap between smORF-FLAG and MitoTracker-Red was quantified, and the colocalisation coefficients for almost all of the Reticular smORFs to Mitochondria was revealed to be very high, showing that these are indeed localised to mitochondrial structures. Since the smORFs tested in this assay were chosen at random from the Poly-Ribo-Seq experiment, it is a significant result that almost half (42.22%) of the tested smORFs are mitochondrial. Several of the mitochondrial proteins required for basic processes in the mitochondria are encoded by mitochondrial DNA (mtDNA) found within the organelle itself. The potential role for nuclear genomic smORFs in the mitochondria may be separate from these primary roles of energy production or perhaps supplement them in some way.

Experimental validation of subcellular localisation was compared to bioinformatics predictions using the TargetP program (Emanuelsson *et al.* 2000; Emanuelsson *et al.* 2007). For all ~800 FlyBase smORFs, only 9.8% are predicted by TargetP to have mTP domains. 17.5% of all the smORFs we found to be translated in S2 cells have mTP motifs, showing that the predicted proportion of mitochondrial SEPs is higher in our target pool. TargetP correctly predicted the subcellular localisation of almost half of the SEPs that I observed to localise to mitochondria. However three SEPs (CG32230, CR30055, CG42497) which have high colocalisation levels to MitoTracker-Red, were given no high-confidence prediction for localisation by TargetP. Additionally, two SEPs

(CG34242 and CG42308) were incorrectly predicted to have a SP motif, but instead localise to the mitochondria. Since TargetP predictions fall short of actual experimental validation, we can assume that the proportion of mitochondria localised SEPs is actually higher in the FlyBase smORFs as well. However even if this number is potentially higher, the overall difference between the FlyBase and Poly-Ribo-Seq smORF pools may be due to the fact that the FlyBase smORFs contain developmentally related smORFs and are therefore more diverse, whilst S2 cells may be preferentially expressing smORFs required for cellular homeostasis. Considering that such a large proportion of S2 cell smORFs are predicted and experimentally verified to associate with the mitochondria, further investigation was conducted into whether they are essential or required for correct mitochondrial function. The role of mitochondria in cells, and the potential cellular function of mitochondrial smORFs, are discussed further in the next chapter.

# Chapter V - Testing smORF function in *Drosophila* S2 cells

## Introduction

This chapter of my thesis describes a series of RNA-interference (RNAi) experiments aimed at exploring the functional characterisation of smORF encoded peptides (SEPs) in *Drosophila* S2 cells. These experiments were conducted to assess whether knocking down smORF gene transcripts can lead to observable cellular phenotypes. If true, this would provide support to our hypothesis that smORFs are indeed important genomic elements. From the experiments discussed in the previous chapter, we obtained corroboration of the translation of a subset of smORFs deemed to be translated by Ribosome Profiling in S2 cells. Those results support the notion that the larger group of hundreds of Poly-Ribo-Seq smORFs are translated into peptides as well. Evidence of smORF translation allowed us to approach the subject of characterising smORF function by using RNA interference (RNAi), with confidence that this method of silencing transcripts would reduce the levels of the SEPs in the cells. RNAi in *Drosophila* elicits a post-transcriptional gene silencing (PTGS) event, specifically of the gene that is chosen to be depleted. In the previous decade, RNAi has been widely used to assess loss-of-function phenotypes of genes within known pathways. RNAi screening is also used to find novel genes that may be involved in particular developmental or cellular processes. Applying RNAi to knockdown smORF transcripts in S2 cells would be the definitive next step in the larger scale characterisation of smORF function at the cellular level.

**Mechanism of Post-transcriptional Gene Silencing**

The PTGS method was first described in transgenic tobacco plants, which were genetically modified to express portions of viral genomic DNA. The expression of this DNA rendered the plant resistant to that virus by RNAi of viral mRNA (Lindbo *et al.* 1993). Later, studies conducted on *C. elegans* identified that the gene silencing response was initiated by the production of double-stranded RNA (dsRNA) molecules that are homologous to the target gene (Fire 1999; Tabara *et al.* 1999). The mechanism was

further unravelled to show that there are two main steps to RNAi, which are initiation, and execution.

dsRNA acts as a template to identify and degrade potential 'non-self RNAs' in order to limit the expression of harmful foreign RNAs, such as those from viruses which hijack the cellular transcriptional machinery to replicate (Vaucheret *et al.* 1998; Fire 1999). dsRNA molecules can be made naturally, being derived from short hairpin RNA's in the genome, called micro-RNAs (miRNA). miRNA help to regulate gene expression and are found within the genome, therefore they are considered a type of endogenous RNAi. RNAi can also be induced by a viral pathogen, or created experimentally and injected into animals, or even added to the growth media of cells in culture. To conduct exogenous RNAi in *Drosophila*, long dsRNAs are made against large stretches of target mRNA, and sequences are processed into smaller dsRNA fragments by Dicer, which is a member of the ATP-dependent RNAse III gene family (Blaszczyk *et al.* 2001). Dicer is a highly conserved RNAse III gene, and homologues have been found in organisms ranging from yeast to humans (Bernstein *et al.* 2001). The resulting small-interfering RNA molecules (siRNA) are between 19-23 nucleotides in length (Blaszczyk *et al.* 2001; Nykänen *et al.* 2001) and have a 2-3 nucleotide overhang at the 3' hydroxyl end.These features of the 'bait' sequence enable the RNAi machinery to detect a bonafide siRNA (Nykänen *et al.* 2001; Kao and Megraw 2004). Mature miRNA and siRNA are very similar in structure, and a pre-miRNA is processed by Dicer in a similar way to dsRNAs.

The execution phase of RNAi begins by the incorporation of the siRNA or miRNA fragments into the 360kDa multiprotein complex called the RNA-induced silencing complex (RISC) (Nykänen *et al.* 2001; Hammond *et al.* 2001). The siRNA is unwound in an ATP-dependent manner, and the single-stranded siRNA acts as a guide for the RISC to target complementary sequences of mRNA in the cell. An endonuclease present in the RISC, called Argonaute, degrades the mRNA, leading to its overall decrease (Liu *et al.* 2003; Hammond *et al.* 2001). The degraded target mRNA can be further recycled into new siRNA and the RNAi process is amplified (Lipardi *et al.* 2001). The schematic presented in Figure 5.1 outlines the RNAi process.

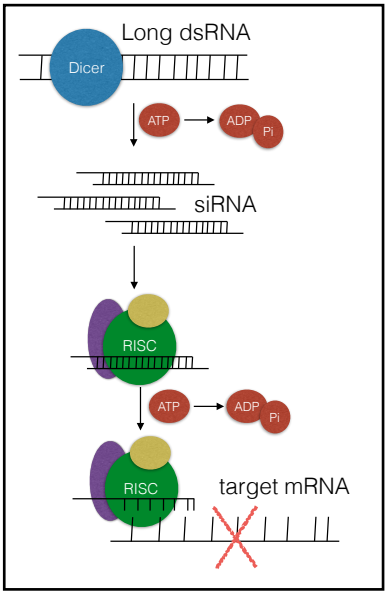



**Figure 5.1: Overview of the mechanism of RNAi in cells using dsRNA**

Exogenous RNA interference is conducted using double-stranded RNA-mediated knockdown. The ~300 bp long duplex is added to the cells, and this is taken up by the cells where in the cytoplasm they are cleaved by the ribonuclease Dicer into ~21nt short interfering RNA's (siRNA's) with 2-3nt overhang at both ends. The 3' of the siRNAs contain a hydroxyl group and the 5' overhang has a phosphate group. The siRNA fragments are incorporated into the RNA-Induced Silencing Complex (RISC). The siRNA is then able to conduct knock-down of complementary target mRNAs by base-pairing via the argonaute endonuclease proteins. This produces an direct decrease in target gene product and indirectly upon the peptide produced from the transcript.

**RNAi Screening in *Drosophila* cells**

RNAi in *Drosophila* cells has now become a mainstream technique used to identify novel genetic hits for further investigation. Traditional RNAi screens are carried out with a specific goal, which typically entails a directed screen for a missing component in a well-characterised pathway, resulting in a catalog of genes involved in a particular process (Boutros and Ahringer 2008). The first RNAi screen in *Drosophila* S2 cells catalogued genes involved in the insulin signal-transduction pathway (Clemens *et al.* 2000). This study showed reliable loss-of-function of target genes simply by adding dsRNA to the media of S2 cells, which are shown to uptake synthetically-made long dsRNAs, and thus show a decrease in target mRNA over time, which can be assessed (by quantifying the peptide product) (Kao and Megraw 2004; Zambon *et al.* 2006). Genome-wide, high-throughput screening has been used to identify candidate genes involved in cell morphology (Kiger *et al.* 2003), cell growth and viability (Boutros *et al.* 2004), protein kinases involved in cell-cycle progression (BettencourtDias *et al.* 2004), and genes associated with cytokinesis (Eggert *et al.* 2004).

The development of a reliable and robust RNAi screen is often a painstaking process. The screen must contain an assay that is specific to the function being investigated. Identifying a hit is aided by some knowledge of the possible phenotypes, or read-outs, that could be expected. These may include up-regulation of certain proteins or reporters, the occurrence of cell death, or the inhibition of the cell cycle. However, the easier it is to screen for a broad and overt phenotype, such as cell lethality, the harder it is to obtain specific functional information of the hits (Boutros and Ahringer 2008). The overall pool of genes being tested will contain a lower proportion of peptides involved in cell death for example, as this is a highly regulated process. There will be proportionally more hits retained when searching for a more subtle phenotype i.e. changes in cell morphology and shape as there could be many potential factors that regulate this process. However, the more specific, technical, and time-consuming an assay is, the more unfeasible it sometimes becomes to conduct on a larger scale.

The use of fluorometric/colorimetric assays, and visual interpretation of results by imaging, have allowed the phenotyping process to become easier. In order to conduct a high-throughput screen of smORF function, we needed to decide upon an observable

cellular or physiological process that smORFs could be involved in, and how best to quantify it. For this purpose, we first attempted the RNAi screen using a cell viability assay with the Alamar Blue reagent, which measures viability of a given cell sample by providing a fluorescent read-out of respiration. In addition to this, using information gained from the previous chapter that described the colocalisation of several S2 cell translated SEPs to the mitochondria, a mitochondrial morphology assay was designed. By using bioinformatic prediction tools such as TargetP, or by extrapolating the results of my colocalisation experiments, we can estimate that anywhere between 9.8% (FlyBase smORFs) and17.5% (Poly-Ribo-Seq smORFs) to 44% (Experimental Data) of SEPs in S2 cells may be localising to the Mitochondria. Localisation of a protein to a specific organelle indicates that there is high likelihood that it performs a function related to that organelle. This encouraged us to search for smORF loss-of-function phenotypes related to the mitochondria.

**Mitochondrial function**

Mitochondria are hypothesised to arise from an α−proteobacterium, which may have been engulfed by progenitors of eukaryotic cells (Lane and Martin 2010). The structure of mitochondria remains similar to that of their Gram-negative bacterial ancestor, with an outer membrane to separate from the cytoplasm, and an inner membrane, which envelops the mitochondrial matrix. The schematic of a mitochondrion shown in Figure 5.2, highlights some of the many processes that are conducted inside mitochondria. Mitochondria are known as the ‘powerhouse of the cell’ as they are the cell’s main site of energy production. The matrix contains the mtDNA and the various enzymes that perform oxidative metabolism. The reason for this is to enable the mitochondrion to retain control over the gene expression of these specific respiratory chain proteins, as it allows each mitochondrion to quickly respond to changes in the membrane potential. There is a direct link between the amount of mtDNA within the mitochondria and cellular respiration rates, and the depletion of mitochondrial genes causes mitochondrial disease (Lane and Martin 2010). The imaging conducted for the previous chapter showed that our S2 cells contain a tightly connected reticular network of mitochondria all over the cell cytoplasm.

Mitochondria perform respiration by using the energy stored in ingested food, primarily through the metabolism of glucose. Glucose is converted into pyruvate in the cytoplasm and transported into the mitochondria. Metabolism occurs in the mitochondrial matrix from the oxidation of glucose and fatty acids into acetyl-CoA. Processing of acetyl-CoA by the Krebs cycle utilises reduction of NAD+ and FAD to yield $CO^2$, the final product of oxidative metabolism. Oxidative phosphorylation takes place on the inner mitochondrial membrane (James *et al.* 2012). Transport of electrons from NADH and FADH2 carrier molecules is converted into energy as a membrane potential. The energy stored in the proton gradient drives the synthesis of ATP (Lodish 2008). Disruptions of the metabolic process in mitochondria can be measured using assays which target key steps in oxidative phosphorylation, and the production of ATP (Mcbride *et al.* 2006; James *et al.* 2012).

It is thought that all of the proteins involved in oxidative phosphorylation are localised to the inner mitochondrial membrane (Ernster and Schatz 1981). The inner membrane is only permeable to pyruvate and fatty acids while excluding large molecules. The outer membrane of mitochondria contains porin channels which allow the diffusion of cytosolic contents smaller than 6 KDa (peptides of about 50 amino acids) into the inter-membrane space (Lodish 2008; Park *et al.* 2010). It is quite interesting that this size range corresponds to that of smORF peptides, and could be a clue as to why such a large proportion of smORFs localise to the mitochondria.

Programmed cell death (PCD), another very important cellular process, is mediated by the mitochondria. Drosophila-inhibitor of apoptosis protein (DIAP1) acts as a ‘brake’ on apoptosis. Apoptosis is promoted by the protein Reaper, which is a 68 aa smORF peptide (White *et al.* 1994). Reaper is part of the well characterised RHG (Reaper, Hid, Grim) protein family, which encode peptides that localise to the outer mitochondrial membrane (Thomenius *et al.* 2011). Expression of Reaper disrupts DIAP1 function by directly binding DIAP1 and translocating it to the mitochondria, where DIAP1 is ubiquinated and destroyed (Freel *et al.* 2008; Steller 2008). Reaper and Hid expression has been shown to induce ‘swelling’ of mitochondria in S2 cells (Abdelwahid *et al.* 2007).

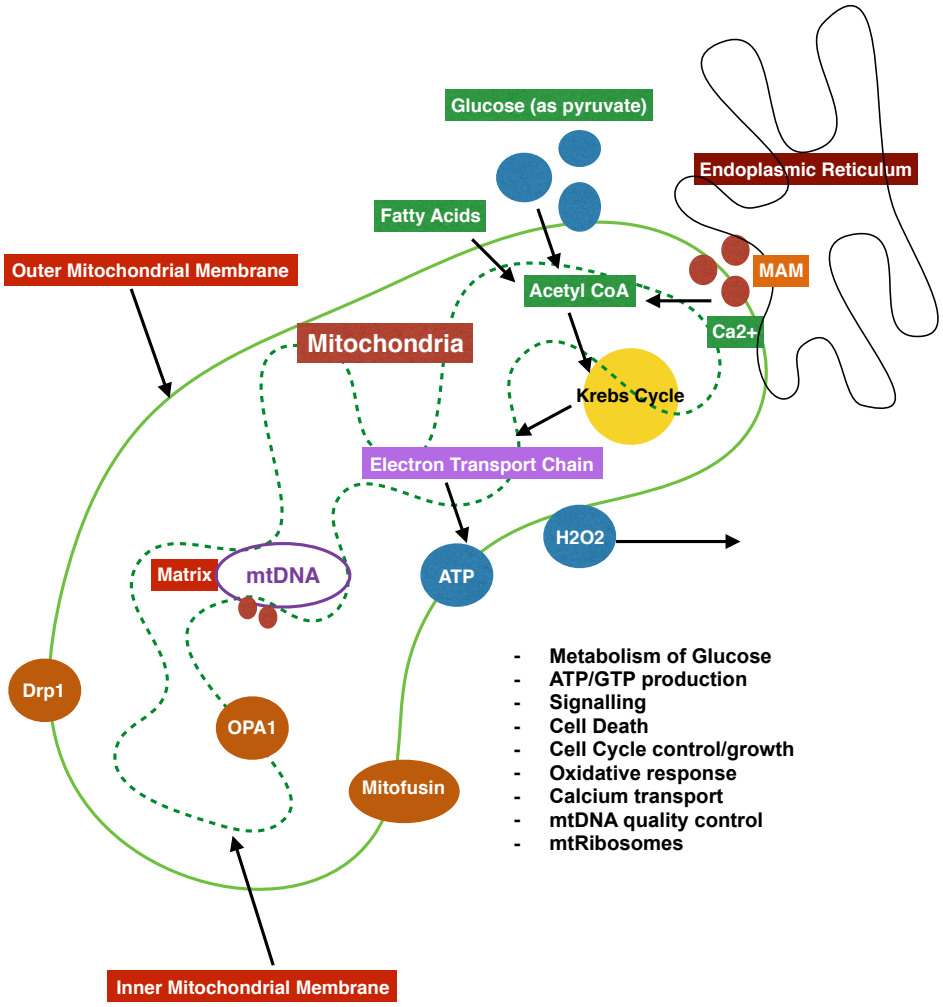



**Figure 5.2: The structure and function of mitochondria in eukaryotic cells** The mitochondria has a matrix within the inner mitochondrial membrane (IMM) and these structures are enveloped within the outer mitochondrial membrane (OMM). The matrix contains the mtDNA, which encodes for some essential peptides of the electron transport chain. Glucose and fatty acids transported across the OMM, are metabolised on the IMM via the Krebs cycle, which generates a membrane potential, allowing the electron transport chain to generate ATP. During oxidative stress, hydrogen peroxide is released from the mitochondria to the nucleus and other organelles. The mitochondrial-associated membranes (MAM) shared with the ER allow influx of Calcium, which is involved in the conversion of Acetyl CoA. Drp1, OPA1 and Mitofusin regulate the fission and fusion of mitochondria to increase mtDNA quality, or respond to cellular energy requirements.

Mitochondrial morphology has been shown to be affected when cells are treated with various stimuli. Developmental cell death of larval tissues during pupation also induces mitochondrial remodelling, causing mitochondria to change from reticular distribution to 'condensed bodies' *in vivo* (Goyal *et al.* 2007). A key phenotype exhibited by cells under stress is the perinuclear aggregation of mitochondria. The cytokine Tumour Necrosis Factor (TNF) is a known promotor of cell death. TNF treatment of the murine fibrosarcoma L929 cell line showed a distinct translocation of mitochondria from a dispersed network into a peri-nuclear cluster (DeVos *et al.* 1998). This spatial change in mitochondrial distribution preceded cell death by several hours. The morphological effect on mitochondria by TNF, was due to the disruption of kinesin motor function, which prevented the mitochondria from traveling along microtubules.

Perinuclear clustering of mitochondria has been observed during programmed cell death in *Drosophila* Nurse cells during mid-oogenesis (Tanner *et al.* 2011). In *Drosophila* S2 cells, perinuclear clustering of mitochondria has been shown after prolonged exposure to paraquat, which is a herbicide linked to Parkinson's Disease, and after genetic over-expression of the Parkin protein (Okatsu *et al.* 2010; Ivatt *et al.* 2014). Knockdown of mitochondrial fission and fusion proteins such as Drp1, also results in perinuclear mitochondrial remodelling in mammalian and *Drosophila* cells (GandreBabbe and van der Bliek 2008; Tanner *et al.* 2011).

The aggregation response of mitochondria has been thought to occur in order to increase the concentration of apoptogenic factors near the nucleus, which seems feasible and logical as it is this signalling event that commits the cell to apoptosis (Pucci 2008). It has also been shown that mitochondria fuse together in cells as a response to starvation (Rambold *et al.* 2015). Autophagy and nutrient stress in mammalian cells forces an increase in the transfer of fatty-acids from lipid-droplet stores to the mitochondria. The boost in oxidation of fatty-acids enables starved cells to continue oxidation and escape cell death. The hyper-fusion and aggregation of mitochondria allows rapid distribution of fatty-acids, and the cells maintain almost equal levels of mitochondrial respiration levels as wild-type cells until committing to cell death (Rambold *et al.* 2015). The hyper-fusion of mitochondria occurs prior to the induction of apoptosis under continued stress (Tondera *et al.* 2009). Together, this information strongly indicates that cellular

stress in the form of autophagy, apoptosis, lack of nutrients or nutrient transport, and the inhibition of mitochondrial transport, causes mitochondria to translocate. The clumping, or hyper-fusion of mitochondria occurs as a cell survival mechanism (Tondera *et al.* 2009).

**RNAi screening for smORF function**

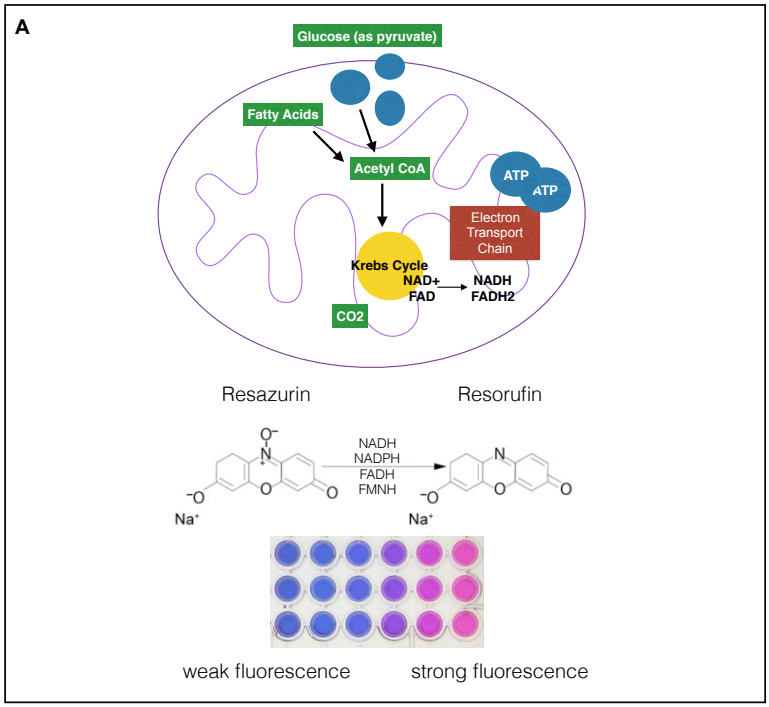
I conducted two different assays on cells after conducting RNAi of target smORF genes. As these assays were conducted in chronological fashion, the best way to understand the information presented here is to compare the assay conducted 'before and after' more information was obtained regarding the localisation of S2 cell smORFs by the tagging-transfection assay described in the previous chapter. Initially, we tried an assay that could be conducted in 96-well plate format, potentially allowing for many samples to be analysed at the same time. The Alamar Blue (AB) reagent can be used to monitor metabolic function and cellular health by reading fluorescence in wells using a plate reader (Rampersad *et al.* 2012). AB, which is a compound called Resazurin, measures the 'reducing environment' of viable cells by acting as an intermediate electron acceptor. As the cells metabolise and grow, AB accepts electrons and changes from a non-fluorescent blue compound into its highly-fluorescent pink/red form, Resorufin. The oxidation-reduction potential of AB allows it to be reduced by NADPH, FADH, FMNH, NADH and the cytochromes of the electron transport chain (Figure 5.3A). The Alamar Blue screen was selected by the lab to be a type of RNAi viability screen, where the result would hopefully reveal which smORFs, are essential to maintain cell viability in the form of normal respiration, as measured by the % of reduction of the Alamar Blue reagent.

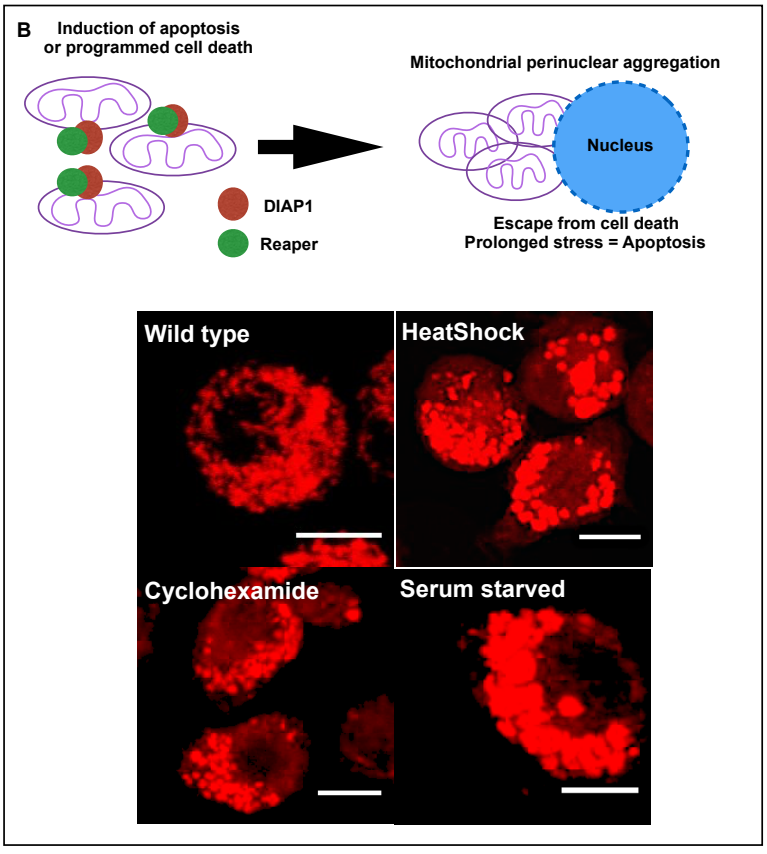
The second study presented in this chapter involves imaging of S2 cell mitochondria after RNAi treatment. These experiments were conducted after complete interpretation of the assessment of the candidate smORFs' translation by Poly-Ribo-Seq as well as the tagging-transfection assay, which also revealed that a large proportion of the smORFs tested actually localise to the mitochondria. I had also previously found that we could observe the mitochondrial stress response of perinuclear aggregation after the induction of cellular stress by various stimuli. The schematic in Figure 5.3B outlines this

mitochondrial stress response and shows examples of a mitochondrial clump/ aggregation phenotype in our S2 cells. By quantifying the proportion of cells that display this mitochondrial stress response in dsRNA treated S2 cells, we would be able to discern which smORFs are inducing stress directly upon the mitochondria, prior to any decrease in respiratory activity or commitment to cell death. Generally, this approach should confer a favourable selection bias on the screen, as we are fishing for phenotypes in mitochondria, where a potentially large proportion of S2 cell smORFs are localised and may be functioning.

**Figure 5.3: The assays used to test smORF function using RNAi**

**A)** Overview of the Alamar Blue assay which is used to assess cellular health by acting as a read-out of cellular metabolism. AB consists of Resazurin reagent, which acts as an intermediate electron acceptor from all the major electron carriers of the mitochondria. If metabolism is taking place, AB is reduced from a weakly fluorescent blue compound (Resazurin) to a highly fluorescent pink compound (Resorufin). The fluorescent signal can then be detected by a Fluorescent plate reader. This assay would allow the rapid screening of several samples at the same time. **B)** The mitochondria imaging RNAi screen, which uses the detection of clumped-mitochondrial morphology as a read-out to detect mitochondrial stress. The schematic shows the proposed model for the induction of mitochondrial stress by modulators of programmed cell death DIAP1 and reaper. In S2 cells, wild type mitochondria appear as strings and small punctae distributed through the cytoplasm in a filamentous network. The induction of cellular stress by Heat-shock, prolonged serum-starvation and incubating cells with the translation inhibitor cyclohexamide leads to perinuclear aggregation and clumping of mitochondria. Scale bars = 5μm.


A
Glucose (as pyruvate)
Fatty Acids
Acetyl CoA
ATP
ATP
Electron Transport Chain
Krebs Cycle
NAD+
FAD
NADH
FADH2
CO2
Resazurin
Resorufin
NADH
NADPH
FADH
FMNH
Na+
Na+
weak fluorescence
strong fluorescence


B
Induction of apoptosis
or programmed cell death
DIAP1
Reaper
Mitochondrial perinuclear aggregation
Nucleus
Escape from cell death
Prolonged stress = Apoptosis
Wild type
HeatShock
Cyclohexamide
Serum starved

# Results

## Section I - Alamar Blue RNAi screen

Alamar Blue reagent is non-toxic and permeable through cell membranes. Fluorescence and absorbance of AB is observed at 560nm excitation and 590nm emission wavelengths. AB reagent is water-soluble and stable in culture medium, though the presence of Foetal Bovine Serum, which is used to supplement Schneider's media for S2 cells, can quench signal. There have not been many studies conducted with AB in *Drosophila* cell lines. The novelty of this approach, and the fact that we had the necessary tools available to conduct the assay, made AB a feasible choice. Cells that are inhibited in growth and metabolism by the presence of a drug or treatment do not effectively reduce the reagent, and show a smaller percentage of AB reduction compared to healthy cells.

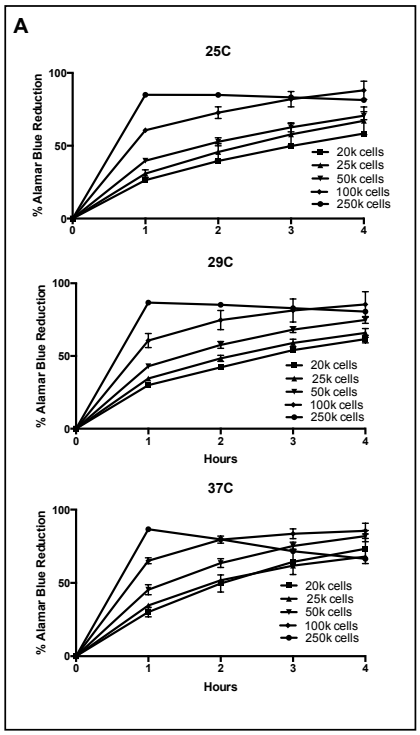
### Temperature Testing of Alamar Blue Assay

The optimum incubation temperature for AB is 37°C, and thus far this assay has mainly been used for mammalian and bacterial cells, for which 37°C is the optimal growth temperature (HamalainenLaanaya and Orloff 2012). The first step to see whether we can use AB to conduct our screen on smORF function would require testing whether we can use it at lower temperatures, as *Drosophila* physiological temperature is 25-29°C, and cells and flies are known to exhibit a heat-shock response at 37°C (Lindquist *et al.* 1984). A previous study on *Drosophila* S2 cells shows that *Pseudomonas* exotoxin causes apoptosis, and this was determined by incubation of *Drosophila* cells with AB at 37°C after treatment with toxin (Sharma and Fitzgerald 2010). In order to clarify whether the results of this study are due to the *Pseudomonas* exotoxin or heat shock, I decided to perform a temperature test on the cells.

S2 cells produce large amounts of Heat-Shock Proteins when kept at a temperature 37 °C for longer than 5 hours (Cevallos and Sarnow 2010). Therefore, I assessed whether we obtain similar AB response curves from cells plated at lower temperatures, such as S2 cells, which require culturing and incubations at 25°C. S2 Cells were grown in flasks resuspended in Serum free media, counted, and diluted to concentrations of $2.0x10^5$, $2.5x10^5$, $5.0x10^5$, $1.0x10^6$ and $2.5x10^6$ cells per mL, of which 100µl was plated in 10

wells each. 10 wells of media-only were also plated in order to obtaining a reading of background signal. Three identical plates were prepared for each temperature that was to be tested. In order to keep conditions consistent with dsRNA treatment, cells were supplemented with 100µl of 10%FBS media after 1h, and then allowed to grow for 48 hours at 25°C. After the 2-day growth period, the media of the cells was aspirated and 200µl of 10% AB (v/v) Serum Free media was added to the cells in order to obtain a reduction curve. Each plate was kept at different incubation temperatures, (25°C, 29°C and 37°C) after the addition of AB, and readings were obtained at 1-hour intervals for 4 hours followed by a final reading at 24 hours. After this time, all the plates were autoclaved at 155°C, and an endpoint reading of AB reduction was obtained and averaged. The assay endpoint is required in order to determine the % of AB reduction, and is constant for all experiments conducted from this point (115322 fluorescent units). The % reduction calculation is conducted by manufacturer's instructions, described below. The results of this experiment are presented in Figure 5.4A.

To conduct an assay under consistent experimental conditions, especially when plating a large number of samples at the same time for comparison, requires optimisation of the technique so that it can be reliably reproduced at a larger scale. Due to this, we had to determine the starting cell density at which to plate the S2 cells, as well as the technique of adding and removing media and reagents from the sample. S2 cells double after every 24 hours, and the amount of dsRNA added to the cells is determined by the cell density at the beginning of the experiment. The following tests were conducted to select the lowest number of cells that we can use for plating, while remaining within the recommended experimental range for conducting the AB assay (under 250,000 cells/ well). The general formula for dsRNA to cell ratio is 1.0µg per 250,000 cells (*Drosophila* RNAi protocol DRSC), and I aimed to be far below this starting point due to limited amounts of dsRNA as well as the fact that AB reduction is hampered at cells grown above this density. A series of dilutions were plated in triplicate wells using a range from $1.0x10^4$ cells/ml (1,000 cells/well) to $5.0x10^5$ cells/ml (50,000 cells/well). Cells were plated using the method described above, this time using a multichannel pipette to aid the scaling up process of the assay, and the plates were incubated at 25°C. Previously, reagents were being pipetted into the wells carefully using a single channel




| B | Trial 1 | Trial 2 | Trial 3 | Trial 4 |
|---|---|---|---|---|
| Conditions | Addition of AB into Growth media | Cells Resuspended in 10% AB SF media | Removal of growth media and addition of 10% AB SF media | Removal of growth media and addition of 10% AB SF media |
| AB incubation | 4h | 4h | 4h | 4h |
| Results | inconsistent | inconsistent | consistent | consistent |

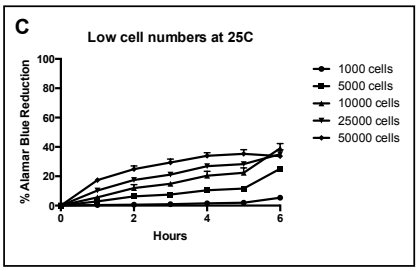



**Figure 5.4: Optimisation of the Alamar Blue assay**

**A)** A range of 20,000 to 250,000 S2 cells were plated in wells of a 96-well plate and incubated with AB reagent at 25°, 29° and 37°C for 4 hours. Fluorescent readings were obtained every hour and the average percent reduction of AB by the cells was calculated. These tests showed that incubating cells at different temperatures does not change the response curve of AB **B)** Different methods used to add AB reagent to the cells. The method of removing growth media followed by the addition of 10% AB reagent in Serum free media gave the most consistent readings. **C)** Test of plating very low cell density in order to conserve dsRNA.

pipette. Though this approach was thought to lead to error, the results from the temperature test showed that there was not too much variation between the different dilutions of cells. As

this was not feasible when dealing with hundreds of samples, different approaches were considered regarding the adding of AB reagent to the cells after the 2 day phase of growth and incubation with dsRNA. The table in Figure 5.4B summarises various techniques that were used to add Alamar Blue into the media of the cells. Fluorescent readings were obtained at 1-hour intervals up to 6-hours after adding AB reagent. The first trial was conducted by adding 20µl of AB into the existing growth media of the cells. The readings obtained for each cell density range were highly variable. The second trial involved the aspiration of the 2-day old growth media, and cells were resuspended in fresh 10% AB serum-free media and incubated for 4 hours. The results from resuspending the cells were even more variable than adding AB into the growth media. The act of resuspending cells by pipetting up and down in the wells may have caused removal of some media and cells.

The third and fourth trials were conducted under identical conditions, by removal of growth media and gentle addition of 10% AB serum free media into the wells. The results of these two experiments were averaged and are presented in Figure 5.4C. From this, I saw that plating between 5000-50,000 cells/well is achievable, and still shows a dynamic range of AB reduction over time. The overall percentage of AB reduction is lower than in the results of similar cell density in the temperature-testing experiments. This is most likely due to the transfer of cells into Multichannel Pipette Reservoirs, which may cause some loss of cells as they get stuck to the reservoir plastic, or sink to the bottom before being pipetted. Therefore I decided to plate 25,000 cells/well for the RNAi experiment as it provided a small enough number of cells to require only 0.1µg of dsRNA per well. The incubation time with AB was selected at 4h after addition of 10% AB Serum Free media, as this provided enough time for the reagent to show a large change in fluorescence for cells plated at lower densities.

**Pilot RNAi screen**

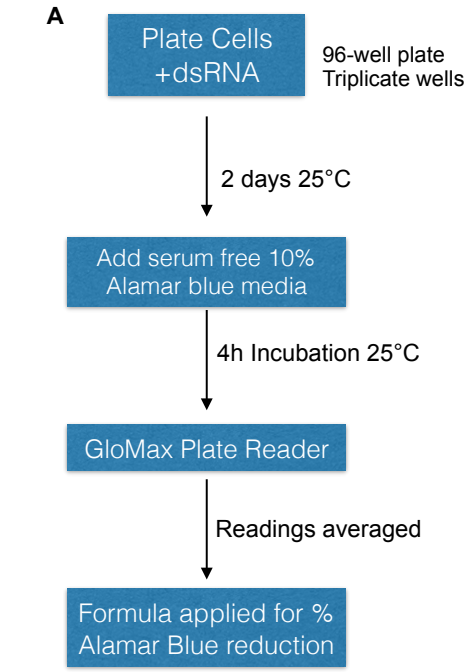
The experimental protocol that was followed for the first RNAi AB Pilot has been shown in Figure 5.5A along with the names of the various candidate smORF and control dsRNAs that were tested using the AB assay in the Pilot screens. The smORFs that were to be tested with the AB assay RNAi screen were selected from the Peptide Atlas database of smORFs (Desiere *et al.* 2006), as the data from the Poly-Ribo-Seq experiments had not yet been analysed. These SEPs have evidence of translation through mass spectrometry studies in the fly. We aimed to identify genes that encode peptides which, when removed, impinge on the levels of mitochondrial metabolism as compared to wild type and positive controls. This can be calculated by by comparing percentage AB reduction of cells by applying the formula provided by the manufacturers. (AB signal reduced by sample - AB signal from media only ÷ 115322 fluorescent units - AB signal from media only).

As previously mentioned, an important role of mitochondria is the regulation of programmed cell death, or apoptosis (Liu *et al.* 1996; Susin *et al.* 1996; Desagher and Martinou 2000). Apoptosis can be initiated externally by cell surface receptors leading to caspase activation, or internally regulated by the mitochondria (Suen *et al.* 2008). *Drosophila* Inhibitor of Apoptosis Protein-1 (DIAP1) is a key player in the 'gas and brake' model of apoptosis in flies (Steller 2008). DIAP1 is essential for cell survival, and this is the reason that this gene was chosen as a positive control for cell death. Elimination of DIAP1 results in early embryonic cell death in *Drosophila* (Wang *et al.* 1999). Treatment of *Drosophila* SL2 cells with DIAP1 dsRNA showed caspase activation and cell death after a decrease of DIAP1 mRNA expression (Zimmermann *et al.* 2002). In *Drosophila*, caspases have been found to be constitutively active, and require the constant presence of DIAP1 binding in order to inhibit the apoptotic response (Zimmermann *et al.* 2002; Freel *et al.* 2008). Another control for cell growth was used, in the form of Cyclin-dependent kinase-1 which is encoded by the *cdc2/Cdk1* gene (Edgar *et al.* 1994) and will be referred to as cdc2 in this thesis. Cdk1/Cdc2 dsRNA has been shown to cause cell death and the formation of giant cells, in a previous study conducted on S2 cells (Björklund *et al.* 2006). It was hoped that either of

these controls would show us the effect on cellular respiration when cell survival proteins are knocked down.

A total of 31 samples were plated in triplicate, including a no-dsRNA/wild-type cell control in a 96-well plate. The Pilot screen was run in triplicate, using the initial cell density of 25,000 cells/well incubated with 0.1μg of dsRNA. The remaining 3 wells of each plate were dedicated to media-only, to provide a background reading, which is subtracted from the fluorescence readings obtained from the test samples. dsRNA-Untreated cells and dsBlueScript-treated cells were used as wild-type controls for these experiments. The dsRNA-untreated control acted as a baseline of S2 cell mitochondrial metabolism, in order to measure any variation imposed by the experimental protocol. dsBlueScript is designed against a bacterial plasmid that is not found in S2 cells and is used as a control to show if there is any variation imposed simply by the addition of a non-specific dsRNA to the cells, which may elicit an RNAi response from the cellular machinery without targeting a specific mRNA. The results of the first round of experiments are shown in Figure 5.5B.

The expected result from the knockdown of DIAP1 would be a lower percentage of AB reduction than that in wild-type (dsRNA untreated) and BlueScript dsRNA treated S2 cells. Though percent reduction of DIAP1 RNAi is significantly different to untreated cells ($p<0.0001$), it is actually higher than % AB reduction in wild type cells, which is the opposite to what was expected. The difference between percent AB reduction of DIAP1 RNAi cells is non-significant compared to dsBluescript treated cells ($p=0.4368$). Percent AB reduction of cells treated with cdc2 dsRNA did not differ significantly when compared to dsBluescript treated cells either ($p=0.0938$). This result rendered the base-line of the experiment flawed, as there was no difference between the positive and negative controls, therefore the difference between DIAP1 and wild type cells cannot be attributed to a specific knockdown of DIAP1 and may just be arising due to the induction of an RNAi response. Without a significant difference between these conditions, we would be unable to interpret results from smORF RNAi samples as cells with decreased viability should show a percentage reduction of AB significantly lower than the negative controls and around the range of positive-control treated cell values.




| Controls | smORF | smORF |
|---|---|---|
| Untransfected | CG34250 | CG14818 |
| BlueScript | CG14701 | CG3321 |
| D-iAP1 | **CG32267** | CG9603 |
| cdc2 | CG40439 | **CG33714** |
| tarsal-less | CG17343 | CG17931 |
| Reaper | CG32069 | CG13551 |
| NAT1 | CG13018 | CG8860 |
| RenLuc | CG33672 | CG32368 |
| | CG2021 | **CG7630** |
| | CG11985 | CG5174 |
| | CG18624 | CG14235 |
| | CG9065 | CG17059 |

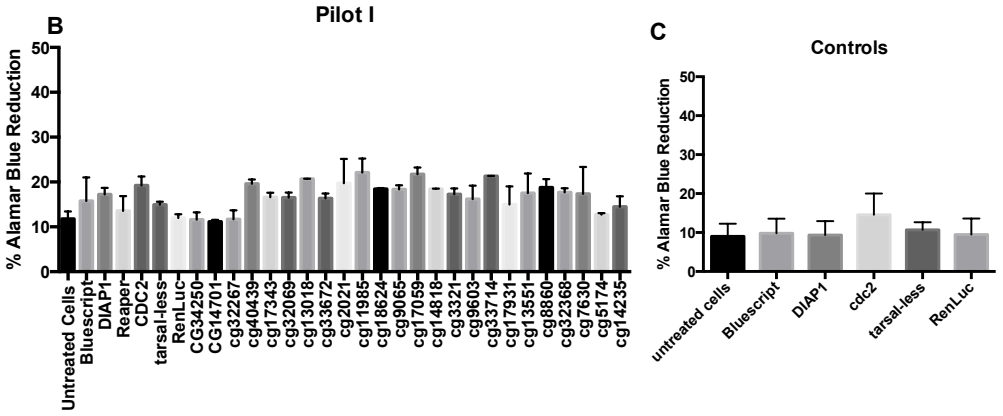



**Figure 5.5: The Alamar Blue Assay to detect smORF function**

**A)** Overview of final AB assay protocol: 25,000 S2 cells were plated in each well of a black 96-well plate. 0.1µg of dsRNA were added to cells in triplicate wells. Plates were sealed and left to grow for 2 days. The growing media of the cells was then replaced by 10% (v/v) AB media and fluorescent readings were obtained after 4 hours at 25 °C. The dsRNAs tested in the AB assay are listed in the table. Tarsal-less, Renilla-Luciferase, and Bluescript dsRNAs are used as negative controls, while DIAP1 and Cdc2 are used as positive controls. smORFs tested in both screens are in bold **B)** Results of the first pilot screen. No significant difference is observed in average %AB reduction between the untreated cells and the controls. **C)** Repeat of the AB assay with just the control duplexes and double the amount of dsRNA, again no significant difference is observed.

This led us to question whether the knock-down is taking place at all, especially for the positive controls. Therefore, I decided to increase the amount of dsRNA in an attempt to verify the affects of the positive controls. Figure 5.5C shows the results of a duplicate experiment where I retested the positive and negative controls. Cells were plated as described above, but this time using double the amount of dsRNA. The percentage reduction of AB in DIAP1 and Bluescript dsRNA treated cells was still nonsignificant ($p=0.8792$). dsRNA treatment targeting cdc2, a cell cycle protein, had a higher level of AB reduction than all the other controls. However, compared to Bluescript treated cells, this result was still nonsignificant ($p=0.1156$). Therefore, the need arose for verifying the knockdown of the positive controls.

**Validation of positive controls of cell death and cell growth**

Upon analysing the results from the first pilot screen, I decided to experimentally verify knock-down of the DIAP1 and cdc2 transcripts using semi-quantitative RT-PCR on dsRNA treated cells. Primers for RT-PCR were designed for targeted amplification of the endogenous cdc2 and DIAP1 transcripts. This was achieved by excluding regions of the transcripts complementary to the dsRNA sequence from the amplicon, thus preventing any signal from dsRNA amplification from distorting the measurements. Samples were prepared using 1mL of cells in serum-free media, plated at a concentration of $2x10^6$ cells/mL. 8µg of each dsRNA (Bluescript, cdc2 and DIAP1) was added to the cells, and after a 1h incubation, the cells were supplemented with 1mL of 10% FBS supplemented media and allowed to grow for 4 days. Cells were harvested and pelleted for RNA extraction using Trizol Reagent. The results of the RT-PCR are shown in Figure 5.6A. The RT-PCRs showed that there is knockdown of cdc2 and DIAP1 transcripts in cells treated with each dsRNA when compared to levels of expression of these genes in Bluescript dsRNA-treated cells. This shows that our dsRNAs are indeed working in reducing the expression of their positive control targets.

I decided to conduct a cell count of cells treated with the control dsRNAs. The results of this experiment are presented in Figure 5.6B. From each of the 2 repeats of the RT-PCR experiment, I sampled 100µl of cells and plated them on acid-treated coverslips. The cells were allowed to adhere for 1h, and fixed with 4% paraformaldehyde, followed by a

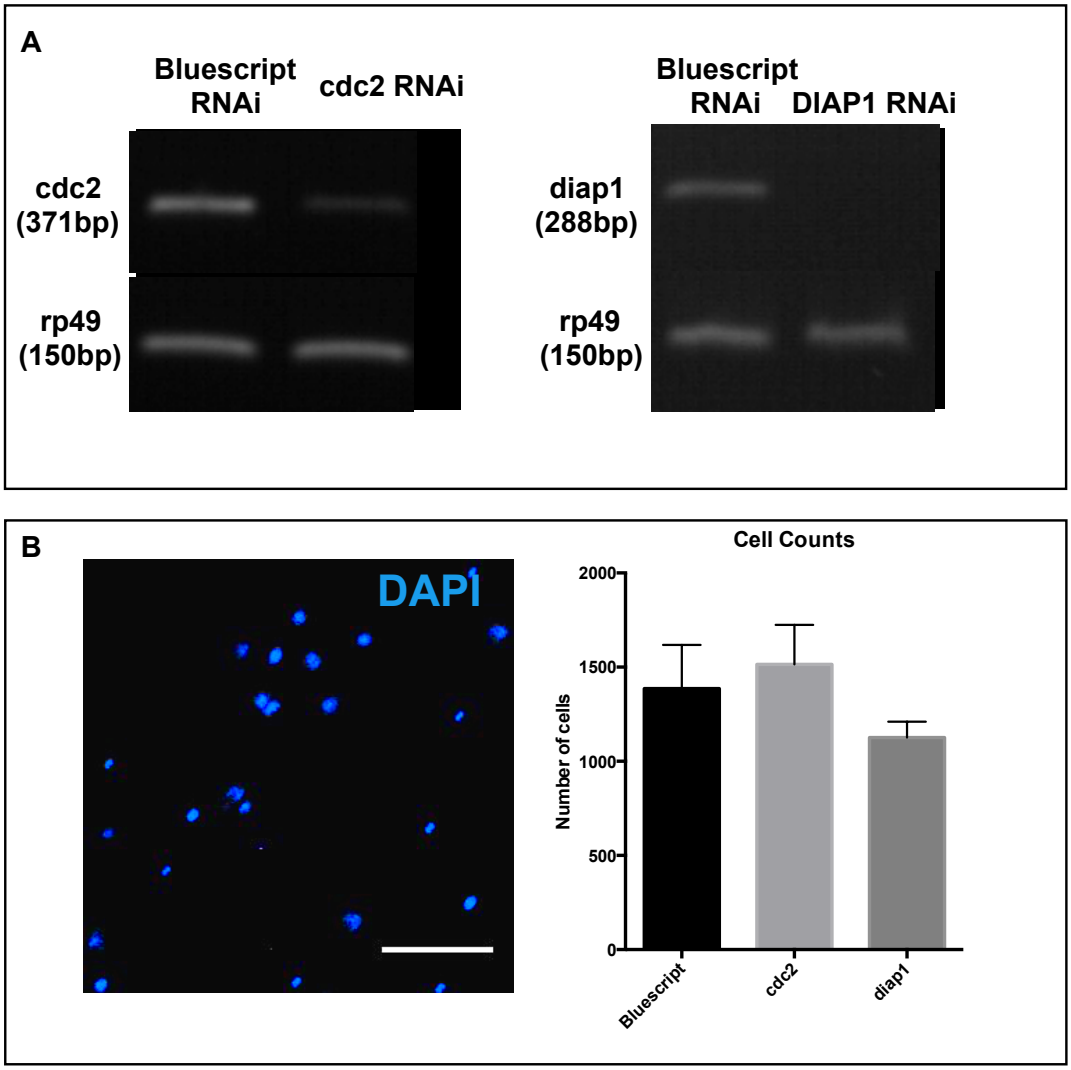



**Figure 5.6: Validating the knock-down of positive controls**

**A)** Semiquantitative RT-PCR to asses level of knockdown of control genes in S2 cells using the dsRNA-soaking method. The gel image shows significant knockdown of both DIAP1 and cdc2 as compared to the expression of these in cells treated with the dsRNA negative control Bluescript. Rp49 is used as a loading control. **B)** A small sample of cells were stained with DAPI and the nuclei were counted to assess cell number. Though there is no significant difference between the test and control samples DIAP1 nuclei numbers are slightly lower than negative control cells. Scale bar = 10μm.

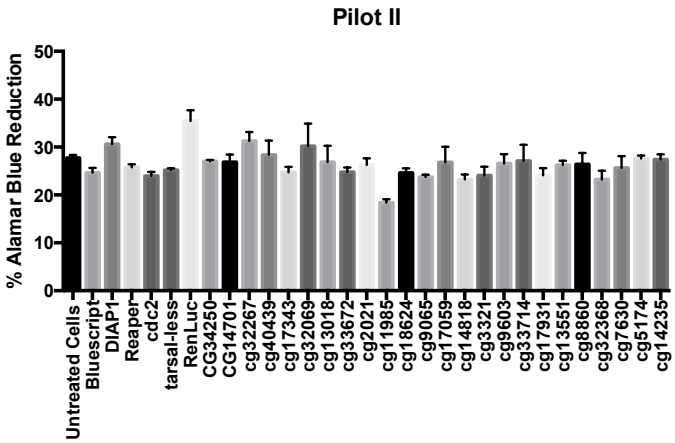
DAPI staining to label nuclei. The coverslips were then mounted on glass slides with a drop of Vectashield and imaged on a wide-field fluorescent microscope (Leica). pictures were taken of each slide, and the nuclei were counted using automated particle recognition in the ImageJ program and the total number of cells counted in each replicate was averaged. Neither cdc2 nor DIAP1 RNAi's showed significantly higher or lower cell numbers than samples treated with Bluescript dsRNA, however the DIAP1 RNAi cell counts were slightly lower. The fact that knockdown of controls can be seen by RT PCR, and cell counts are slightly lowered in DIAP1 dsRNA samples was encouraging enough to try the assay again with a larger amount of dsRNA.

**Optimisation of Alamar Blue RNAi screen**

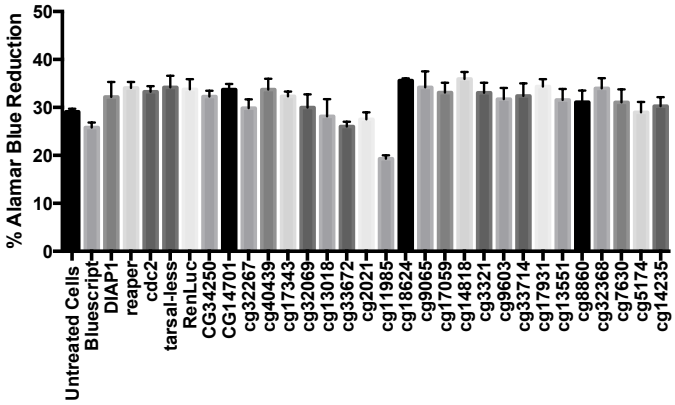
The Pilot RNAi was repeated again, at the same starting cell density as before, but in the hope of seeing an improved effect of dsRNA treatment, I added twice the amount of dsRNA (0.2µg) in both these attempts and Pilot 2 cells were incubated for 2-days and Pilot 3 cells were allowed to grow for 4-days, to help improve knock-down and see a more pronounced effect of RNAi on the signal derived from AB. The results of these rounds of experiments are shown in Figure 5.7A,B. The overall percentage of AB reduction has increased in these experiments compared to the first pilot. This increased signal is probably due to the longer 4-day growth phase and therefore the presence of a higher number of cells .

Unfortunately, the results from these two experiments showed a wild-type (Bluescript-dsRNA treated) level of percentage AB reduction in positive-control DIAP1 and cdc2 dsRNA treated cells. This indicated that knockdown of these genes over the course of both of 2 and 4 days does not impose a decrease in AB signal as expected; as neither DIAP1 or cdc2 RNAi show a decreased level of AB reduction. We have previously verified the effectiveness of these control dsRNAs, so this means that the effect of knocking down DIAP1 and cdc2 does not decrease the levels of respiration in the cells. We have previously verified the effectiveness of these control dsRNAs, so this means that the effect of knocking down DIAP1 and cdc2 does not decrease the levels of respiration in the cells. Although the experimental conditions were identical in both of

A




B




C

| ANOVA summary | | | |
|---|---|---|---|
| | Pilot 1 | Pilot 2 | Pilot 3 |
| F | 14.28 | 21.84 | 25.60 |
| P value | <0.0001 **** | <0.0001 **** | <0.0001 **** |
| differences among means statistically significant? | Yes | Yes | Yes |
| R square | 0.6334 | 0.7254 | 0.7559 |

**Figure 5.7: Alamar Blue Assay results after optimisation of RNAi**

**A)** Results of the second pilot screen, which was conducted with double the amount of dsRNA as the first pilot. There are no significant effects on AB reduction observed from positive controls **B)** Results of the third pilot screen, conducted with double the dsRNA and 4-days of incubation of dsRNA and cells before the AB assay with similar lack of effect from positive controls as Pilots 1 and 2. **C)** ANOVA summary of all three pilot screens, each pilot screen results are statistically significant ($p<0.0001$).

these Pilot runs, the percentage AB reduction between the two experiments for specific dsRNAs was variable and inconsistent. One-way ANOVA (Figure 5.7C) of all three screens shows that we can reject the null hypothesis, that there is no effect of RNAi, as the F ratio of each Pilot run is much higher than 1. Each experiment also has a highly significant P-value, therefore we can conclude that the groups are not all the same. Without significantly higher or lower percentage AB reduction of positive controls, there is not much hope of extracting meaningful data by comparing each smORF RNAi with wild type conditions. "Dunnett's multiple comparisons test" was used to compile the multiple t-tests of each smORF and control RNAi treatments against Bluescript-dsRNA treated cells.

The table in Figure 5.8 shows the statistical significance of percentage AB reduction for each dsRNA in each experiment compared to that of Bluescript-dsRNA treated cells. The chart depicts the number of times a particular hit or control was significantly different to the negative control. The graph at the bottom of Figure 5.8 summarises the hits that were highly significant (either $P \leq 0.001$ or $P \leq 0.0001$). The reason for only considering samples at high confidence levels was that untreated cells show a significant ($P \leq 0.05$) difference to Bluescript-dsRNA treated cells in some of the tests. There are four samples which stand out in these screens, one of them being DIAP1 (2 experiments) and another is RenLuc, which is a negative control targeting the Renilla-Luciferase transcript, which is not endogenously expressed in S2 cells.

Two smORF RNAi samples stand out, and these are CG11985 and CG17059. CG11985 has been described to be a component of the Splicing factor 3B (SF3B) complex by sequence similarity. The SF3B complex associates with the splicing factor SF3A complex and a 12S RNA unit to form the U2 small nuclear ribo-nucleoproteins complex and may be involved in mitotic spindle assembly (Flybase). The SEP from CG11985 has not been studied individually or in depth and could be a good candidate for further characterisation as lowering the expression of CG11985 resulted in decreased viability of S2 cells in 2 experiments. The peptide derived from CG17059 has been described to be a homolog of Myeloma-overexpressed 2-like (MYEOV2-like) protein (FlyBase). Myeov2 is a 57aa peptide, but is annotated as 252aa in mammals. *Myeov2* encodes a

| | Pilot 1 | Pilot 2 | Pilot 3 |
|---|---|---|---|
| Untreated Cells | * | * | * |
| DIAP1 | ns | **** | **** |
| Reaper | ns | ns | **** |
| CDC2 | ns | ns | **** |
| tarsal-less | ns | ns | **** |
| RenLuc | * | **** | **** |
| CG34250 | * | ns | **** |
| CG14701 | ** | ns | **** |
| CG32267 | * | **** | ** |
| CG40439 | * | ** | **** |
| CG17343 | ns | ns | **** |
| CG32069 | ns | **** | *** |
| CG13018 | ** | ns | ns |
| CG33627 | ns | ns | ns |
| CG2021 | * | ns | ns |
| CG11985 | **** | **** | **** |
| CG18624 | ns | ns | **** |
| CG9065 | ns | ns | **** |
| CG17059 | **** | ns | **** |
| CG14818 | ns | ns | **** |
| CG3321 | ns | ns | **** |
| CG9603 | ns | ns | **** |
| CG33714 | *** | ns | **** |
| CG17931 | ns | ns | **** |
| CG13551 | ns | ns | **** |
| CG8860 | ns | ns | **** |
| CG32368 | ns | ns | **** |
| CG7630 | ns | ns | **** |
| CG5174 | ns | * | * |
| CG14235 | ns | * | *** |

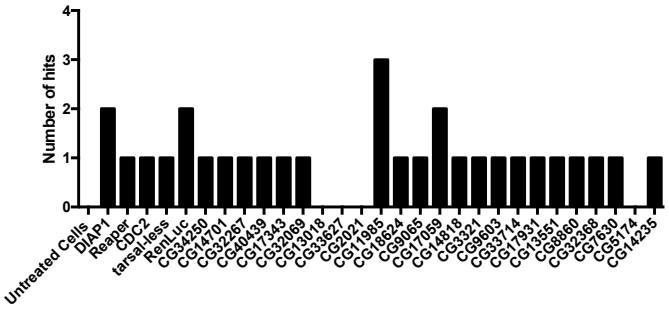



**Figure 5.8: Summary of the three Alamar Blue pilot screens**

'Dunnett's multiple comparisons t-test for each dsRNA tested in the AB assay. Each column shows whether the %AB reduction is significantly different to the Bluescript dsRNA control for each Pilot each experiment (****P<0.0001,***P<0.001,**P<0.01, *P<0.05, Non-significant, ns P>0.05). The graph at the bottom shows the number of times each dsRNA shows up as statistically significant compared to Bluescript in each pilot. There are two smORFs (CG11985 and CG17059) that stand out from this analysis.

peptide which has been implicated to play a role in regulation of the cell cycle (Ebina *et al.* 2013). Since the positive controls of cell cycle and cell death did not seem to show a significant effect in this screen, it is difficult to say whether these two smORFs are actually producing an effect because if they are, it is stronger than the metabolic response of cells in which DIAP1 and cdc2 knock-down is occurring.

**Section II - smORF RNAi screen in S2 cell Mitochondria**

Although the Alamar Blue screen was not as successful as we had hoped, there were several observations made through those experiments that enabled us to move forward with the RNAi screening technique. The validation of RNAi knockdown of the positive controls for cell viability (DIAP1 and cdc2) by RT-PCR showed that the dsRNA soaking technique is effective at inducing knock-down by this mechanism. The Poly-Ribo-Seq experiments also revealed which smORFs are being translated at high levels our lab's S2 cell line, which is important because every S2 cell population is unique due to the rapid accumulation of mutations based on the growth and cell culture conditions used to maintain the cells.

It had been previously validated that MitoTracker-Red dye specifically labelled mitochondria in live cells (Chapter IV). S2 cell mitochondria were observed to be a mixture of puncta and strings, displaying a snapshot of the dynamic nature of mitochondria in healthy cells. The stringy and punctate mitochondria are distributed symmetrically around the cell cytoplasm in a filamentous network. When S2 cells were heat-shocked, serum starved for prolonged periods, or treated with cyclohexamide, the mitochondria became aggregated into large punctae, and localised asymmetrically in the cell (Figure 5.3B). As described in the introduction of this chapter, mitochondria have been shown to remodel in response to stress. Since we had preliminary evidence of this phenotype existing in our S2 cell population, I decided to conduct an imaging screen of mitochondrial morphology after RNAi treatment in order to identify smORFs with mitochondrial function. This mitochondrial stress response has been described in the literature by many, but it has never specifically been used as a direct indicator of cellular stress after inducing gene loss-of-function.

**smORF RNAi and imaging S2 cell mitochondria**

Cultured S2 cells were resuspended in serum-free media and counted using a haemocytometer before dilution to 250,000 cells/mL. Cells were plated at a density of 250,000 cells per well, which is a standard number for imaging. 6-well culture plates were prepared to contain an acid-treated coverslip upon which 2µg of dsRNA (or in untreated cells, 2µl of milliQ H2O) was added. 1mL of cells were added to the coverslip and allowed to mix with the dsRNA by gentle swirling of the plate. After a 45 minute incubation, the cells + dsRNA were supplemented with an equal volume of Schneider's media containing 20% FBS and allowed to grow for 2 days (48 hours) on the coverslip. After this time, MitoTracker-Red, diluted to 500nM in 10% FBS Schneider's media, was added to the wells and incubated for 1h. After which the media was removed and the cells were fixed with 4% paraformaldehyde, stained with DAPI for 10 minutes to label the nuclei. The coverslip was mounted on glass slides containing a drop of Vectashield before Imaging.

In order to remove sampling bias depending on the dsRNA being used, the samples were randomised during the coverslip mounting process. Slides were labeled with a random letter and coordinate of the well in which they were plated. This information was recorded by another member of the lab who kept it hidden from me until after I had completed the imaging and analysis. This enabled me to score the mitochondrial morphology phenotypes using a blind method, in which there was no distinction made between the slides for wild-type controls, positive controls and test smORFs. After scoring the imaged cells for aggregated mitochondria, the proportion of aggregated cells was calculated from the total cells imaged. If a significant proportion of cells in a smORF RNAi sample displayed perinuclear aggregation of mitochondria when compared to wild-type and negative control samples, it would indicate that knocking down of the smORF gene product causes mitochondrial stress. Figure 5.9A summarises the process of plating and randomisation of the samples for analysis.

There were two possibilities for conducting the imaging of S2 cell mitochondria in this experiment; the first would be to image the S2 cells using basic wide-field fluorescence, which is faster and less expensive than the second option of Laser Scanning Microscopy. LSM confocal microscopy is most often used for more detailed imaging of

samples, and if the phenotype could be reliably discerned from wide-field fluorescence microscopy, this would be the preferred method as it would enable us to apply this study to a larger number of samples. As shown in Figure 5.9B-C, samples imaged using wide-field microscopy resulted in signal from mitochondria lying above and below the plane of focus, which lowers the resolution of the image. Blind scoring of mitochondria using these low resolution images was inconsistent when repeated for the same sample, as it was often difficult to differentiated true 'clump' morphology (Figure 5.9C) from ubiquitously distributed mitochondria (Figure 5.9B) even though they were discernible by eye.

LSM imaging proved to be a more successful approach for this assay as Z-stack images were obtained of each plane of focus through the cells at intervals of 0.2µm. The Z-stack projections of these images showed the mitochondria in higher detail, and distinct differences between normal (Figure 5.9D) and clumped or aggregated (Figure 5.9E') mitochondrial morphologies could be observed. In order to make imaging of the samples more efficient, single-channel (Red) imaging was conducted to speed up the collection of between 50-100 cells per repeat of each dsRNA. Z-stack imaging not only provides clearer mitochondrial morphologies, but images can be projected into a 3D reconstruction and rotated around the Y axis using the 3D projection tool in the ImageJ program. Figure 5.9F shows 3D projections of S2 cells at 0°, 45° and 90° rotation of the the axis. By scoring the cells using 3D rotation of the images, it was even easier to differentiate between the normal (Ubiquitous) and abnormal (Clumped) mitochondrial classes. Aggregation of mitochondria can occur along the Z-plane, which cannot be seen without imaging the entire cell from the coverslip to the slide. 'Ubiquitous' mitochondria were described as a reticular distribution of mitochondrial structures evenly in the cell, along each plane. Clumped mitochondria have large gaps in the cell, and most of the mitochondria are polarised towards one side of the cell, along any plane. For robustness, only the severely clumped mitochondria, where all the mitochondria of the cell are contained in only one half of the cell, were quantified as abnormal.

For the Mitochondrial morphology screen, dsRNAs were designed to target all transcripts of the smORF gene, therefore for 8 SEP's found in dicistronic transcripts,

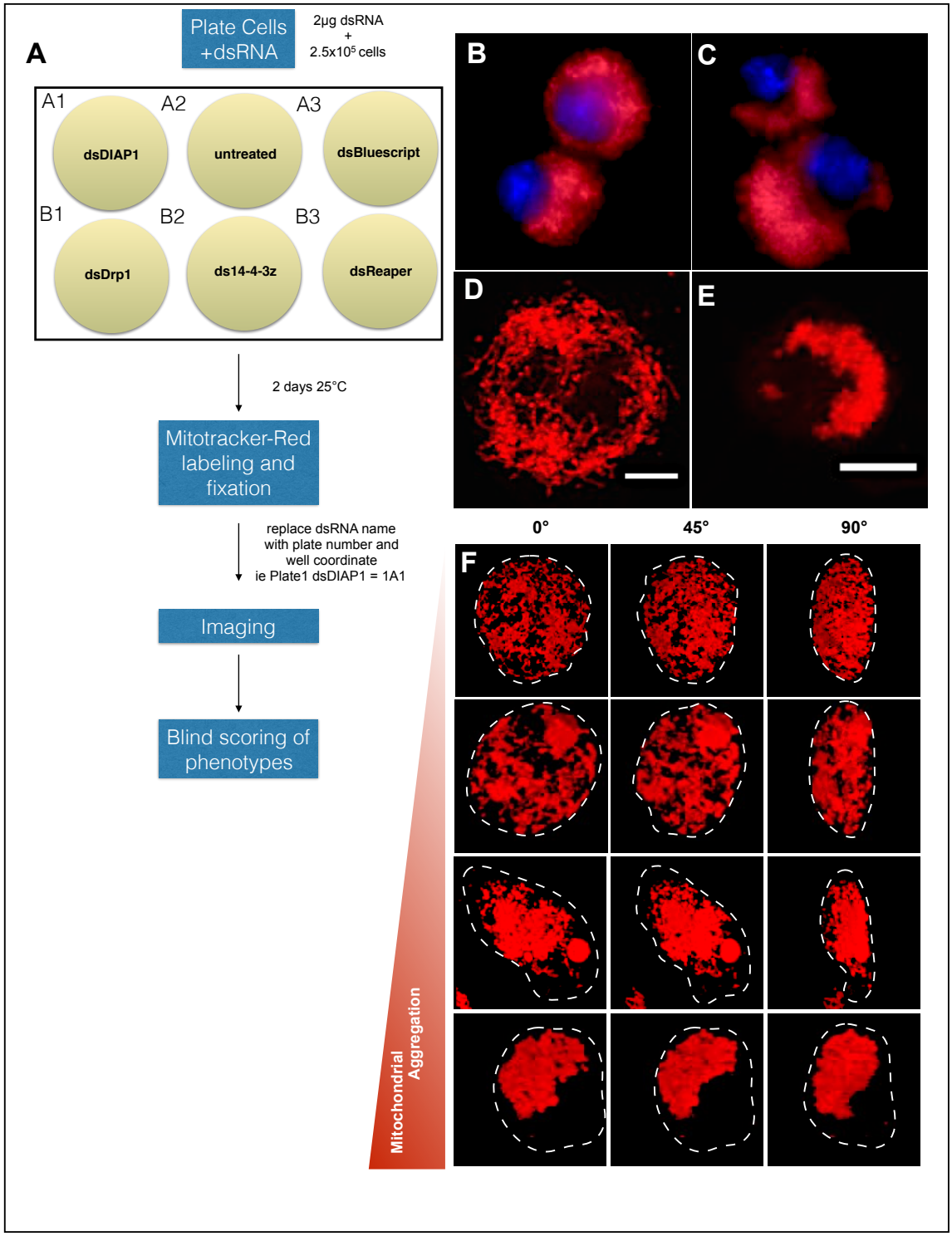
there are four dsRNAs targeting their transcripts for knockdown. For this reason, uORFs were excluded from this screen as the dsRNA would also knock-down the canonical downstream ORF, and the effect observed would not be specific to the smORF being tested. The pilot list of this screen included all the smORFs that we tested in the tagging-transfection assay. Figure 5.10 summarises the list of smORFs and controls that were tested in this screen including 27 smORFs, 4 dicistronic smORFs and 6 lncRNA smORFs, of which 3 are monocistronic and 4 are polycistronic.

**Results from mitochondrial morphology RNAi screen**

Cells were prepared for imaging as described for the controls. The samples contained 250,000 cells at the time of plating, which is the standard protocol for the plating of cells on a coverslip. Due to the semi-adherent nature of S2 cells, some of these would be inevitably lost through staining and fixing, but this is still a very high number of cells to image at high magnification using the Laser Scanning technique. Therefore, the number of cells imaged was limited to a realistic range which could be imaged for hundreds of samples. A minimum of 50 cells and up to a maximum of 100 cells were imaged per replicate of each smORF. The images obtained were named according to the randomised coordinate of the sample, and blind-scored from 3D reconstructions. After scoring the cells, the proportion of imaged cells with abnormal mitochondria was determined for each repeat of every sample. For each round of repeats, wild-type and positive controls were also present amongst the sample wells and hence repeated several times. Each smORF RNAi sample was tested at a minimum in triplicate. The results of the mitochondrial morphology RNAi Screen are shown in Figure 5.10A, as well as the complete list of different smORFs and controls that were targeted for RNAi (Figure 5.10B). The One-way ANOVA of the results (Figure 5.10C) showed that the differences between the average percentages of cells with abnormal mitochondria between each sample in the overall experiment were statistically significant ($P< 0.0001$). The results from this experiment are depicted in '% abnormal mitochondria', after averaging the proportions of abnormal mitochondria cells in the imaged samples. dsRNA-Untreated cells and dsBlueScript-treated cells were used as wild-type controls for these

**Figure 5.9: smORF RNAi and imaging S2 cell mitochondria**

**A)** Overview of the mitochondrial morphology assay using RNAi against DIAP1, reaper, Drp1 and 14-3-3 as positive controls. Bluescript-RNAi and untreated cells are used as negative controls. 250,000 S2 cells are plated onto a coverslip in the presence of 2 μg of dsRNA and allowed to grow for 2 days; the cells are then labelled with MitoTracker Red and fixed before imaging and blind scoring of phenotypes. **B)** Imaging normal mitochondria using wide-field microscopy. **C)** Imaging clumped mitochondria after RNAi using wide-field microscopy. This was conducted at the highest magnification of this microscope (100x, Oil Immersion Lens), however the low resolution of the microscope means that it is difficult to resolve the filamentous network of mitochondria in B from the clumped mitochondria in C. **D)** Normal cells imaged using LSM confocal microscopy increases resolution, allowing visualisation of stringy and punctate mitochondria **E)** LSM imaging of clumped mitochondrial morphology can be seen as different to the normal mitochondria in D. **F)** 3-D reconstructions of Z-stacks taken using LSM, the three columns of images show the mitochondrial morphology at different angles (0, 45 and 90 degrees on y-axis) while severity of mitochondrial clumping increase from top to bottom. Scale bars = 5μm.


A
Plate Cells +dsRNA
2µg dsRNA + 2.5x10⁵ cells
A1
dsDIAP1
A2
untreated
A3
dsBluescript
B1
dsDrp1
B2
ds14-4-3z
B3
dsReaper
2 days 25°C
Mitotracker-Red labeling and fixation
replace dsRNA name with plate number and well coordinate ie Plate1 dsDIAP1 = 1A1
Imaging
Blind scoring of phenotypes
B
C
D
E
0°
45°
90°
F
Mitochondrial Aggregation

experiments. dsRNA-untreated cells had 16.76% (±4.2, 5 repeats) and dsBluescript treated cells 21.18% (±8.2, 7 repeats) of cells with aggregated mitochondria. I also designed a dsRNA against tarsal-less, which is not expressed in S2 cells. dsRNA-tal would act as a negative control for whether RNAi against an untranscribed smORF causes any changes to mitochondrial morphology. dsRNA-tal cells contained an average of 29.68% (±7.72, 3 repeats, $P>0.05$, ns) of cells showing aggregated mitochondria, which is similar to the wild type and dsBlueScript-treated results.

I also used several positive control dsRNAs based on their effects in published studies, in which they are shown to be implicated in cellular stress, death, or mitochondrial re-arrangements. DIAP1 was once again used as a positive control for cell death, as in the AB screen. In the mitochondrial morphology screen, DIAP1 dsRNA treated cells had an average of 41.67% (±7.46, 6 repeats, $P\leq0.0001$) cells with abnormal mitochondria. Dynamic-Related protein 1 (Drp1) regulates mitochondrial fission by directly interacting with fission and fusion molecules found on the outer mitochondrial membrane. It has been well characterised that knocking down Drp1 causes perinuclear aggregation of mitochondria in many cell types (GandreBabbe and van der Bliek 2008). In our screen, RNAi against Drp1 showed 62.61% (±11.90, 5 repeats, $P\leq0.0001$) of cells with aggregated mitochondria, which is one of the most severe phenotypes characterised in this assay.

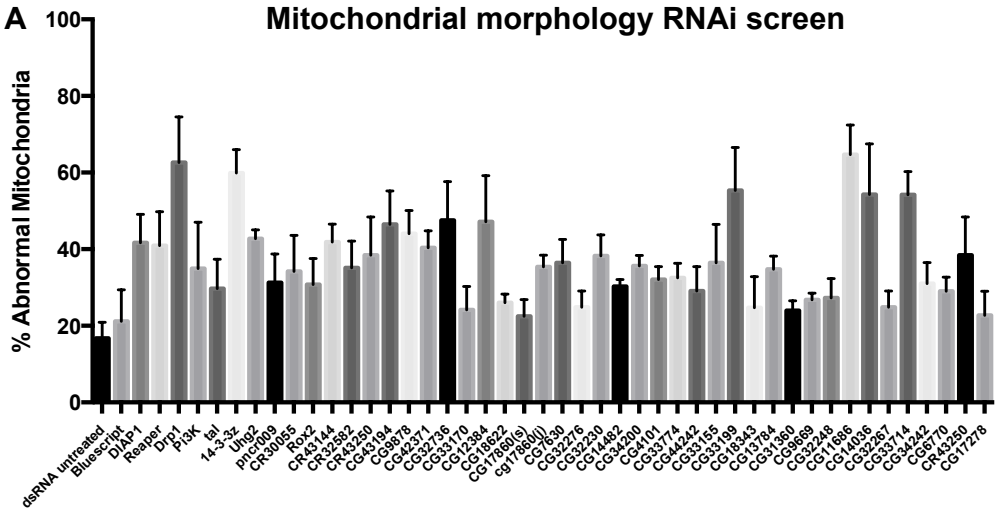
This screen also tested whether the effect of knocking down other known genes would have an effect on mitochondrial morphology. PI3K68D is an enzyme which regulates endosome and phagosome maturation, as well as intracellular trafficking of lysosomes (Rusten *et al.* 2004; Velichkova *et al.* 2010). Cells treated with dsRNA against PI3K68D showed 34.93% (±12.11, 3 repeats, $P>0.05$, ns) of the cells with abnormal mitochondria, which was a nonsignificant result compared to wild type cells. Reaper is a well characterised inducer of cell death (Freel *et al.* 2008; Thomenius *et al.* 2011), but it was previously unknown what effect RNAi against Reaper has on mitochondrial morphology. Cells treated with Reaper dsRNA showed 40.94% (± 8.82, 4 repeats, $P\leq0.01$) of abnormal mitochondria, which is almost as severe as the phenotype of DIAP1. RNAi against the master cellular regulator 14-3-3z at 59.88% (±6.1, 3 repeats $P\leq0.0001$) resulted in almost as many abnormal cells per sample as Drp1. 14-3-3zeta is

involved in many vital cellular functions, including correct chromosomal separation during mitosis (Su *et al.* 2001). The knockdown of this protein has been shown to induce apoptosis in hepatic cancer cells (Lee *et al.* 2014).

RNAi mediated knockdown of these known important cellular components helps determine a potential range of abnormal mitochondrial populations that can be expected in the cell populations. As opposed to the results from the Alamar Blue screen, the negative control Bluescript-dsRNA treated sample is not significantly different to the wild-type dsRNA-untreated cells. Additionally, the negative control for smORFs, tarsal-less, was also non-significant to both of the wild-type controls. DIAP1, Reaper and Drp1 positive control samples show a significantly higher proportion of abnormal mitochondria than wild-type cells, as does the 14-3-3z dsRNA. Knocking down the known lncRNA Uhg2 (Tycowski and Steitz 2001) causes significant mitochondrial stress (42.71±2.35% cells with abnormal mitochondria, 3 repeats, $P \leq 0.01$). Uhg2 (U-snoRNA host gene) ORFs themselves are not translated (Tycowski and Steitz 2001), the knocking down of snoRNA transcripts can have a deleterious effect on cells because it would target the translational machinery of the cell.

In all, the smORF RNAi screen showed that 41% (15/37) of the total smORFs knocked down using RNAi, have an effect on mitochondrial morphology. 80% (12/15) of the smORFs showing a RNAi phenotype have either experimental data or TargetP prediction of localisation to the mitochondria, thus showing that this screen is a direct assessment of the function of mitochondria associated SEPs. The complete results have been summarised in a table (Figure 5.11) showing for each individual smORF the average percentage of abnormal mitochondria, standard deviation, number of repeats and statistical significance against dsRNA-BlueScript treated cells. The smORFs were scored by a lettering system according to known information about localisation, either through experimental validation or by a TargetP mTP prediction (Figure 5.11B). The results discussed below are smORFs which have a phenotype in this screen, as well as an indication of mitochondrial localisation. These smORFs, which have been given a score of A*, A, or B and are classed as real hits in the mitochondrial morphology RNAi screen, will be referred to as Functional mito-SEPs.

A




B

| Controls | smORF | smORF | Polycistronic smORFs | lncRNA |
|---|---|---|---|---|
| Untransfected | CG32230 | CG9669 | CG15386/CG42371 | CR32582 |
| BlueScript | CG15456 | CG13784 | CG42497/Tim10 CG9878 | CR43144 |
| tarsal-less | CG33199 | CG17680 | CG32736/CG42308 | CR43250 |
| Pi3K | CG33155 | CG4101 | CG43210/CG43194 | pncr009:3L (ORF1-4) |
| Reaper | CG33774 | CG18622 | | CR30055 (ORF1,3,4) |
| D-iAP1 | CG14482 | CG32248 | | Rox2 (ORF1+ ORF2) |
| Drp1 | CG12384 | CG18343 | | |
| 14-3-3z | CG33714 | CG31360 | | |
| *Uhg2* | CG6770 | CG14036 | | |
| | CG44242 | CG33170 | | |
| | CG7630 | CG32267 | | |
| | CG34200 | CG334242 | | |
| | CG32276 | CG14104 | | |
| | CG17278 | | | |

C

| ANOVA summary | |
|---|---|
| F | 8.899 |
| P value | <0.0001 **** |
| differences among means statistically significant? | Yes |
| R square | 0.7792 |

**Figure 5.10: Complete results from mitochondrial morphology RNAi screen**

**A)** This graph shows the results from the mitochondrial imaging RNAi screen. The proportion of cells containing abnormal, clumped mitochondria were averaged over all the repeats for each dsRNA and are shown on the y-axis. Positive controls in this screen showed a significant difference to wild type and negative controls. Statistical analysis of these results is presented in Figure 5.11 **B)** Table showing the list of duplexes tested in the mitochondrial morphology RNAi screen. The columns show the controls, monocistronic, dicistronic and long non-coding RNA encoded smORFs tested in this screen. CG7630 and CG32230 are present in both AB and mitochondrial morphology screens **C)** ANOVA summary of the RNAi screen data showing the high statistical significance of the results within the screen.

All four dicistronic smORF genes tested in this experiment also elicit a mitochondrial stress response, three of these(CG42497/CG9878, CG42371/CG15386, CG32736/ CG42308) have mitochondrial localisation, and the results from the screen show that knocking down these smORFs has a highly significant effect. Of these, CG42497/ CG9878 has now been annotated as Tim10 on the basis of sequence homology (Flybase). The Tim9-Tim10 chaperone complex is responsible for the transfer of hydrophobic precursor proteins through the mitochondrial inter-membrane space, where it is localised (Schmidt *et al.* 2010). Interestingly, RNAi against Hemotin (CG43194/ CG43210) also causes mitochondrial abnormality (46.46±8.75%, 4 repeats, $P \leq 0.0001$), even though Hemotin peptides do not localise to mitochondria in my colocalisation experiments (Chapter IV). TargetP predicts the Hemotin-ORF2 peptide to contain a mTP sequence with high confidence. Additionally, Stannin, the vertebrate homolog of Hemotin has been shown to co-localise to mitochondrial markers, so a possibility does exist that there is some role for Hemotin in the regulation of mitochondrial dynamics (Davidson *et al.* 2004).

The monocistronic mitochondrial smORFs CG7630, CG32230, CG33199, CG14036, CG14014, and CG33714 are all positive hits in the screen. CG7630 has a mild phenotype of abnormal mitochondria (36.42±6.15%, 4 repeats, $P \leq 0.05$) and I have found it to be co-localised to MitoTracker-Red. CG7630 has also been detected in the proteome of lipid droplets (Cermelli *et al.* 2006). Mitochondria utilise oxidation to enzymatically reduce Fatty-acids into metabolic intermediates that drive respiration (Rambold *et al.* 2015). Fatty acids are contained within lipid droplets, which are transported into the mitochondria from the Mitochondria Associate Membranes of the ER (Vance *et al.* 1997). The mitochondrial morphology screen shows that removing the Lipid droplet CG7630 by RNAi has an effect on mitochondria which is similar to when cells are put under starvation conditions (Rambold *et al.* 2015). CG32230 has a similar effect on mitochondrial morphology as CG7630, (38.24±5.52%, 3 repeats, $P \leq 0.05$). CG32230 is described as a NADH dehydrogenase subunit (complex 1) and this is inferred from sequence or structural similarity of CG32230 to this peptide in human cells. NADH is one of the main electron carrier molecules responsible for the conversion of energy in the mitochondria, and is localised to the inner mitochondrial

**Figure 5.11:Summary of mitochondrial morphology screen results**

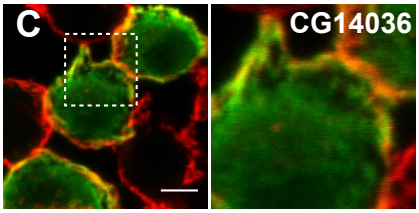
**A)** Summary table of the mitochondrial morphology screen showing the average percentage abnormal mitochondria, standard deviation, number of repeats and statistical significance of each test dsRNA result against dsRNA-BlueScript treated cells (Dunnett's multiple comparison t-test). (****P<0.0001,***P<0.001,**P<0.01, *P<0.05, Non-significant, ns P>0.05). The results of the screen are scored by a lettering system according as shown in the table in panel B. **B)** Table describing the various score classes of hits in the screen by known information about localisation either by experimental validation or a TargetP mTP prediction and phenotype in the screen **C)** Cortical distribution of CG14036 in the sub-cellular distribution study, where some signal can also be observed in cytoplasmic puncta. Scale bars = 5μm.

**A**

| dsRNA | % abnormal | ± | # of Cells | # of Repeats | significant | Phenotype | Score |
|---|---|---|---|---|---|---|---|
| Bluescript | 21.18 | 8.18 | 380 | 7 | | | |
| untreated | 16.76 | 4.18 | 196 | 5 | ns | | |
| **DIAP1** | **41.67** | **7.46** | **436** | **6** | **** | **Yes** | |
| **Reaper** | **40.94** | **8.82** | **222** | **4** | ** | **Yes** | |
| **Drp1** | **62.61** | **11.90** | **289** | **5** | **** | **Yes** | |
| Pi3K | 34.93 | 12.11 | 168 | 3 | ns | no | |
| tal | 29.68 | 7.72 | 221 | 3 | ns | no | |
| **14-3-3z** | **59.88** | **6.10** | **156** | **3** | **** | **Yes** | |
| **Uhg2** | **42.71** | **2.35** | **204** | **3** | ** | **Yes** | |
| pncr009 | 31.25 | 7.52 | 186 | 3 | ns | no | |
| CR30055 | 34.18 | 9.41 | 218 | 3 | ns | no | **F** |
| Rox2 | 30.75 | 6.81 | 243 | 4 | ns | no | |
| **CR43144** | **41.88** | **4.65** | **153** | **3** | ** | **Yes** | **A** |
| CR32582 | 35.14 | 6.99 | 223 | 4 | ns | no | **D** |
| **CR43250** | **38.36** | **9.99** | **141** | **3** | * | **Yes** | **A** |
| **CG43194 CG43210** | **46.46** | **8.75** | **259** | **4** | **** | **Yes** | **B** |
| **CG42497 CG9878** | **44.07** | **6.02** | **194** | **3** | *** | **Yes** | **A*** |
| **CG42371 CG15386** | **40.38** | **4.40** | **157** | **3** | ** | **Yes** | **A*** |
| **CG32736 CG42308** | **47.54** | **10.07** | **213** | **3** | **** | **Yes** | **A** |
| CG33170 | 24.18 | 6.06 | 225 | 4 | ns | no | |
| **CG12384** | **47.18** | **12.00** | **213** | **4** | **** | **Yes** | **C** |
| CG18622 | 26.00 | 2.27 | 355 | 3 | ns | no | |
| CG17680(s) | 22.45 | 4.40 | 201 | 3 | ns | no | **D** |
| cg17680(j) | 35.35 | 3.06 | 167 | 3 | ns | no | |
| **CG7630** | **36.42** | **6.15** | **196** | **4** | * | **Yes** | **A*** |
| **CG14014** | **42.92** | **8.34** | **221** | **3** | ** | **Yes** | **A** |
| CG32276 | 24.86 | 4.23 | 175 | 4 | ns | no | **D** |
| **CG32230** | **38.24** | **5.52** | **159** | **3** | * | **Yes** | **A** |
| CG14482 | 30.25 | 1.83 | 240 | 4 | ns | no | **D** |
| **CG34200** | **35.64** | **2.76** | **214** | **4** | * | **Yes** | **C** |
| CG4101 | 32.05 | 3.40 | 232 | 3 | ns | no | |
| CG33774 | 32.50 | 3.84 | 175 | 3 | ns | no | |
| CG44242 | 29.06 | 6.38 | 206 | 3 | ns | no | **D** |
| CG33155 | 36.45 | 10.01 | 208 | 3 | ns | no | |
| **CG33199** | **55.34** | **11.24** | **156** | **3** | **** | **Yes** | **A*** |
| CG18343 | 24.71 | 8.09 | 278 | 4 | ns | no | |
| CG13784 | 34.78 | 3.41 | 205 | 3 | ns | no | |
| CG31360 | 23.98 | 2.55 | 232 | 3 | ns | no | |
| CG9669 | 26.76 | 1.77 | 215 | 3 | ns | no | |
| CG32248 | 27.30 | 5.05 | 261 | 3 | ns | no | |
| **CG11686** | **64.69** | **7.70** | **286** | **3** | **** | **Yes** | **C** |
| **CG14036** | **54.26** | **13.22** | **167** | **3** | **** | **Yes** | **A** |
| CG32267 | 24.86 | 4.23 | 151 | 3 | ns | no | |
| **CG33714** | **54.21** | **6.03** | **213** | **3** | **** | **Yes** | **A*** |
| CG34242 | 31.03 | 5.47 | 166 | 3 | ns | no | **F** |
| CG6770 | 29.01 | 3.67 | 172 | 3 | ns | no | |
| CG17278 | 22.71 | 6.30 | 188 | 3 | ns | no | **F** |

**B**

| Mitochondrial Phenotype | MitoTracker Colocalisation | mTP | Score | |
|---|---|---|---|---|
| X | X | X | A* | Positive |
| X | X | | A | Positive |
| X | | X | B | Positive |
| X | | | C | False Positive |
| | X | X | D | False Negative |
| | X | | F | False Negative |



membrane (IMM). Some of the subunits of NADH are translated from mtDNA transcripts (Burman *et al.* 2014), but CG32230 is a nuclear gene encoded peptide (Sardiello 2003). It would be very interesting to investigate whether CG32230 in involved with the effect on neurodegeneration as can be seen for mutants of NADH complex 1 in flies (Burman *et al.* 2014). RNAi against CG33199 shows a very strong phenotype (55.34±11.24%, 3 repeats, P≤0.0001) and there are no previously known phenotypes or biological processes associated with this transcript or the peptide that it encodes. This makes CG33199 a particularly interesting hit to follow up, as it is completely uncharacterised and may be playing a significant role in the mitochondria. RNAi targeting CG14036 also has a very strong phenotype, 54.26±13.22%, 3 repeats, P≤0.0001). Although the subcellular distribution of the CG14036 peptide appears as cortical in the subcellular distribution study, the peptide shows significant overlap to MitoTracker-Red at a level above the cut-offs. Upon re-examination of the images, the CG14036 peptide appears as a stronger Cortical staining, but can also be found in a reticular distribution in the cytoplasm(Figure 5.11C). CG14036 is annotated to have a function in copper-ion homeostasis (Norgate 2007), and there is evidence in the literature that it may be directly related to mitochondrial function in a Parkinson's disease associated study (Cox and Spradling 2009).

CG33714 is annotated as a Steroid receptor RNA activator (SRA) Stem-loop-interacting RNA-binding protein (SLIRP) encoding gene, which encodes a peptide of 90aa in *Drosophila*. The CG33714 SEP is a homolog of SLIRP found in human breast cancer cells (Hatchell *et al.* 2006), and this gene encodes a highly conserved, 109aa peptide in mammalian cells. This peptide is localised to the mitochondria, but has been shown to function as a regulator of hormone-responsive gene expression. CG14014 (42.92±8.34%, 3 repeats, P≤0.01) is annotated in FlyBase to be have 'protein dimerisation activity' and was identified by sequence or structural homology to Basic region-leucine zipper (B-ZIP) transcription factors. These peptides dimerise and bind to sequence-specific DNA to regulate gene expression in response to cellular signals (Fassler *et al.* 2002).

RNAi against the translated lncRNAs encoding mitochondrial SEPs, CR43144 (41.88 ±4.65%, 3 repeats, $P \leq 0.01$) and CR43250 (38.36±9.99, 3 repeats, $P \leq 0.05$) showed a

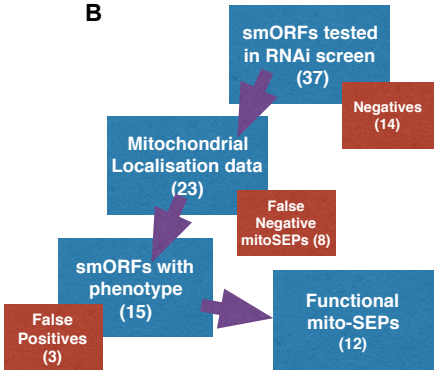
significant effect on mitochondrial morphology. The functions of CR43144 and CR43250 are completely unknown, and the fact that these lncRNA smORFs encode for functional peptides is a very promising result that brings some functional insight: to the extremely challenging task of discerning between 'real' lncRNA genes and mis-annotated smORF genes. There are two smORF dsRNAs (CG12384 and CG34200) that display a positive phenotype in this screen, but were not localised to the mitochondria. CG12384 (47.18±12.00%, 4 repeats,$P \leq 0.0001$) has a strong phenotype but there are no indications that this homolog of Death-Associated-Protein localises to the mitochondria. Programmed cell death and apoptosis occurs on mitochondrial membranes, so it can be argued that CG12384 is actually a positive in the screen, but for robustness of the assay, it is has been considered a false positive result. CG34200 (35.64±2.76%, 4 repeats, $P \leq 0.05$) has a marginally significant phenotype, and no indication of mitochondrial localisation or function and therefore can be considered a true false positive in the screen. It could also be that the dsRNA designed against CG34200 exhibited off-target effects, though no off target effects were predicted for any of the dsRNAs that we designed for this experiment (Snapdragon dsRNA design tool, www.flyrnai.org). There is one last hit, CG11686 that shows a very strong effect on mitochondrial morphology (64.69±7.70, 3 repeats, $P \leq 0.0001$), but in the absence of any localisation or functional data, CG11686 has been classified as a false-positive in this screen until further information is available regarding the subcellular localisation of the peptide. Overall, the mitochondrial morphology RNAi screen is successful at detecting a mitochondrial phenotype from the knock-down of a mitochondria localised smORF peptide with a 52% detection rate. According to the scoring system developed to differentiate functional mito-SEPs from false-positives and false negatives, the screen has a false positive rate of 13% (Figure 5.12A). Figure 5.12B gives an overview of the pipeline used to identify the 12/37 functional mito-SEPs in the RNAi mitochondrial screen, according to their phenotype and localisation. The functional mito-SEPs identified in this screen are summarised, along with a description in Figure 5.12C.

The false negatives from this screen are the 8 mitochondrial SEPs that do not show a phenotype (Figure 5.13A). Three of these have annotated mitochondrial functions, such as the cytochrome b-c1 complex subunit (CG14482), and mitochondrial-ribosomal sub-units CG33155 and CG44242. Since dsRNA treatment against these mitochondrial

A

| A* | A | B | C | D | F |
|---|---|---|---|---|---|
| 5 | 6 | 1 | 3 | 5 | 3 |
| 52% | | | 13% | 35% | |
| Positive mitochondrial hits in the screen | | | False Positives | False Negatives | |

B




C

| Gene ID | aa length | TMMHM | Description | Subcellular Distribution | Subcellular localisation | GO Term |
|---|---|---|---|---|---|---|
| CR43144 | 91 | 0 | lncRNA smORF | Reticular | Mitochondrial | unknown function |
| CR43250 | 73 | 0 | lncRNA smORF | Reticular | Mitochondrial | unknown function |
| CG43194 CG43210 | 88 | 1 | Polycistronic smORF | Cortical | F-actin/ Endosomes mTP (TargetP) | unknown function |
| | 59 | 1 | Polycistronic smORF | Cortical | | unknown function |
| CG42497 CG9878 (Tim10) | 65 | 0 | Polycistronic smORF | Reticular | Mitochondrial | Mitochondrial import |
| | 92 | 1 | Polycistronic smORF | Other Cytoplasmic | | inner membrane translocase (Tim10) |
| CG15386 CG42371 | 77 | 0 | Polycistronic smORF | Other Cytoplasmic | Mitochondrial | unknown function |
| | 87 | 0 | Polycistronic smORF | Reticular | | unknown function |
| CG32736 CG43208 | 79 | 1 | Polycistronic smORF | Reticular | Mitochondrial | unknown function |
| | 61 | 1 | Polycistronic smORF | Reticular | | unknown function |
| CG7630 | 90 | 1 | Monocistronic smORF | Reticular | Mitochondrial | Lipid droplet proteome |
| CG32230 | 83 | 1 | Monocistronic smORF | Reticular | Mitochondrial | NADH dehydrogenase MLRQ subunit |
| CG33199 | 79 | 0 | Monocistronic smORF | Reticular | Mitochondrial | unknown function |
| CG14036 | 93 | 0 | Monocistronic smORF | Cortical | Mitochondrial | copper ion homeostasis |
| CG14014 | 68 | 0 | Monocistronic smORF | Other Cytoplasmic | Mitochondrial | protein dimerisation activity |
| CG33714 | 90 | 0 | Monocistronic smORF | Reticular | Mitochondrial | stem-loop-interacting RNA-binding protein |
| CG12384 | 96 | 0 | Monocistronic smORF | Cortical | F-actin | Death-Associated protein1 (DAP1) |
| CG34200 | 52 | 0 | Monocistronic smORF | Limited | Unknown | unknown function |
| CG11686 | 70 | 1 | Monocistronic smORF | not tagged | Unknown | unknown function |

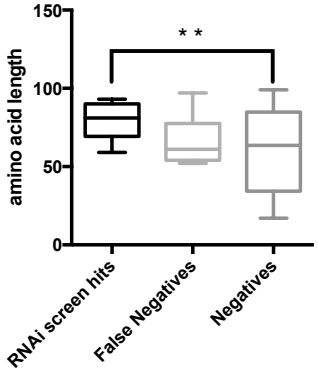
**Figure 5.12: Analysis of mitochondrial morphology RNAi screen results A)** The number of smORFs in each result category, smORFs in categories A*-B (12) are considered High-confidence positive hits as they localise to the mitochondria. C-class results (3) are False-Positives as they show a RNAi phenotype but the SEP does not localise to mitochondria. D&F class results (8) are considered False-Negatives as they are mito-SEPs but do not display a mitochondrial phenotype after dsRNA treatment. **B)** Flow chart showing the pipeline used to categorise Functional mito-SEPs. Boxes in red show the numbers that were discarded at each step **C)** Summary table showing the SEP length, alpha helix prediction, smORF class, subcellular distribution, localisation and any GO terms for the 12 high confidence hits and 3 False-Positive hits.

smORFs did not cause a mitochondrial stress response, it implies that either the RNAi duplex is inefficient or that there is redundancy between these mitochondrial SEP's and other peptides. The latter does not seem to be the case with CG14482, which is a subunit of cytochrome c, an essential component of the electron transport chain; if it were being effectively knocked-down, it would most likely show an effect in the mitochondrial morphology screen. It may also be the case that the mitochondrial SEPs which did not show an effect in the mitochondrial morphology screen, are not needed for normal mitochondrial function but their requirement is heightened under certain conditions such as stress. Another possibility is that the effect of knocking down mitochondrial ribosomal sub-units CG33155 and CG44242 may take a longer time to manifest. Analysis of amino acid lengths between the different groups (RNAi Screen hits, False positives and False negatives) shows that there is no significant difference (P=0.3128) between the functional mito-SEPs (Average=79.69 ± 2.876aa) and False negatives (mitochondrial localised SEP's with no phenotype (Average=66.38±5.487aa) in regards to amino acid length. There is a significant difference ($P \leq 0.01$) between the amino acid lengths of the functional mito-SEPs and the amino acid lengths of the screen 'Negatives' (Average=58.81±7.102aa), which indicates the robustness of this screen as it shows there is a distinct difference between the set of smORFs which are functional, or should be, and those that are not. Whether or not a SEP contains a predicted transmembrane alpha helix does not seem to suggest mitochondrial function, as only 40% of the functional mito-SEPs have transmembrane alpha helix prediction as compared to 63% in the 'real negative' SEPs. The subcellular distributions of all the Real negatives in the screen are either Other cytoplasmic, Limited, or unknown with the exception of CG33774 which was found to localise to mitochondria. A few of them also contain Signal Peptide motif predictions from TargetP (Figure 5.13C). The clear-cut differences between hits and non-hits in this screen indicates that the Mitochondrial morphology RNAi screen is a reliable assay for detecting the function of peptides that are involved specifically in mitochondrial function.

**A**

| Gene ID | aa length | TMMHM | Description | Subcellular Distribution | Subcellular localisation | GO Term |
|---|---|---|---|---|---|---|
| CR30055 | 2-53, 4-17, 5-56 | - | lncRNA smORF | Reticular + Limited | Mitochondrial | unknown function |
| CR32582 | 52 | 0 | lncRNA smORF | Reticular | Mitochondrial | unknown function |
| CG17680 | 97 | 1 | Monocistronic smORF | Unknown | mTP (TargetP) | Positive regulation of immune response |
| CG32276 | 64 | 1 | Monocistronic smORF | Reticular | Mitochondrial | Stress-associated ER protein |
| CG14482 | 57 | 0 | Monocistronic smORF | Reticular | Mitochondrial | cytochrome b-c1 complex subunit |
| CG44242 | 70 | 0 | Monocistronic smORF | Reticular | Mitochondrial | mitochondrial Ribosomal protein S36 |
| CG34242 | 58 | 1 | Monocistronic smORF | Reticular | Mitochondrial | unknown function |
| CG17278 | 80 | 0 | Monocistronic smORF | Reticular | Mitochondrial | unknown function |

**B**




**C**

| Negatives in RNAi screen | aa length | TMMHM | Subcellular Distribution | Subcellular Localisation | GO Term |
|---|---|---|---|---|---|
| pncr009 | 22/30/3 | - | Other Cytoplasmic + Limited | Unknown | unknown function |
| Rox2 | 38/17 | - | Limited | Unknown | unknown function |
| CG18622 | 94 | 1 | Unknown | Unknown | unknown function |
| CG33170 | 71 | 0 | Reticular | Unknown | unknown function |
| CG4101 | 87 | 1 | Unknown | Unknown | unknown function |
| CG33774 | 40 | 1 | Other Cytoplasmic | Mitochondrial | unknown function |
| CG33155 | 60 | 1 | Limited | Unknown | mitochondrial Ribosomal protein L53 |
| CG18343 | 99 | 2 | Unknown | Unknown | unknown function |
| CG13784 | 67 | 2 | Unknown | SP(TargetP) | unknown function |
| CG31360 | 98 | 1 | Unknown | Unknown | unknown function |
| CG9669 | 78 | 2 | Unknown | SP(TargetP) | unknown function |
| CG32248 | 82 | 1 | Unknown | Unknown | unknown function |
| CG32267 | 49 | 1 | Other Cytoplasmic | Unknown | unknown function |
| CG6770 | 69 | 0 | Cortical | F-actin | Nuclear phosphoprotein p8, response to oxidative stress |

**Figure 5.13: False-Negatives and Negatives in the mitochondrial morphology RNAi screen** **A)** Summary table showing different properties of the False-Negative smORFs in the RNAi screen showing SEP length, alpha helix prediction, smORF class, subcellular distribution, localisation and any GO terms. **B)** SEP length analysis of Functional mito-SEPs (RNAi screen hits), False-Negatives and Negative classes. There is a statistically significant difference between the peptide lengths of Functional mito-SEPs (79.69aa ±2.876) and Negatives (58.81aa ±7.102) ($P<0.01$), and no significant difference between Functional mito-SEPs and False-Negative (66.38aa ±5.487) classes. **C)** Summary table showing the properties of the Negatives in this screen along with SEP length, alpha helix prediction, smORF class, subcellular distribution, localisation and any GO terms group of genes.

## Discussion

The Poly-Ribo-Seq experiments conducted by our lab were performed in the *Drosophila* S2 cell line, which is derived from late-stage *Drosophila* embryos (Schneider 1972). The catalog of Poly-Ribo-Seq smORFs translated in our S2 cells gave us a candidate pool of genes to target using RNAi, as we had already observed that these smORF mRNAs undergo active translation. The independent proof of translation provided in Chapter IV also corroborates that these smORF genes are producing peptide products, which can localise specifically to distinct organelles. The use of RNAi in S2 cells, which is a method of Post-transcriptional gene silencing, could potentially be used to discern smORF function on a genome-wide level, since we now have a better understanding of smORF translation at this scale.

The previous chapter of this thesis described several smORF peptides being localised to the mitochondria. Mitochondria themselves contain a small circular genome containing mitochondrial DNA (mtDNA) and much of this mitochondrial genome encodes peptides from genes which are the size of SEPs (Lane and Martin 2010; Nunnari *et al.* 2012). A mammalian mitochondrion can contain over 1500 proteins, expressed in a tissue dependent manner. 13 of those proteins are encoded by mtDNA, which has reduced in size over the course of evolution due to gene transfer events to the host-cell nucleus. Therefore the maintenance and function of the mitochondria are heavily dependent on nuclear-encoded proteins, which are actively shuttled into mitochondrial compartments from the cytosol. Nuclear and mtDNA derived protein subunits are combined into macromolecular complexes, which perform the various roles of a mitochondrion (Schmidt *et al.* 2010).

*Drosophila* mtDNA encodes 13 polypeptides involved in oxidative phosphorylation, 22 tRNAs, and two ribosomal subunits (Garesse 1988; Torres *et al.* 2009;); therefore much of the mitochondrial proteome is encoded by nuclear genes in the fruit fly. Neurological and cardiac diseases, as well as metabolic syndrome, obesity and cancers can be linked to mitochondrial disfunction and have been investigated using the fruit fly as a model (Nunnari *et al.* 2012). In this chapter are described the results of two different RNAi screens that I conducted which led to the discovery of a unique way to screen for smORFs encoding functional peptides that may be playing a role in the mitochondria.

**The mitochondrial morphology screen detects mito-SEP function**

Mitochondria are dynamic organelles that are constantly undergoing division and fusion. Most major regulators of mitochondrial dynamics are conserved from flies to humans. Cell death pathways are regulated through the mitochondrial fission/fusion machinery. By imaging mitochondria in S2 cells, a classification system was made to describe healthy and sick cells based upon the morphologies of the mitochondria. It has been well characterised that mitochondrial aggregation precedes a cell's commitment to apoptosis (Haga *et al.* 2003; Haga *et al.* 2005) by treatment of cells with various apoptotic stimuli. Mitochondrial aggregation has also been attributed to genetic mutation of mitochondrial fission and fusion machinery (GandreBabbe and van der Bliek 2008; Ishihara *et al.* 2003; Qian *et al.* 2012). Interestingly, neurodegenerative diseases can be associated with mitochondrial translocation phenotypes. Parkinson's neurodegenerative disease is known to be associated with mitochondrial defects of perinuclear clustering and there has been a race to characterise the other proteins that may be interacting with the *parkin* disease locus (Greene *et al.* 2003; Park 2009; Okatsu *et al.* 2010; Grünewald *et al.* 2010; DeCastro *et al.* 2011; Wang *et al.* 2012; Ivatt *et al.* 2014; Norris *et al.* 2015).

I decided to observe the effect on S2 cell mitochondrial morphology upon RNAi treatment of a small subset of smORFs from the list of S2 cell translated smORFs generated by the Poly-Ribo-Seq data. The majority of the pool of smORFs selected for this experiment had already been tested in the tagging-transfection assay, so we already had an indication of where these peptides may be functioning. We designed dsRNAs against several smORFs as well as various controls of cell death, and regulators of mitochondrial morphology. In this screen, I was able to establish the functionality of several smORFs by imaging RNAi treated S2 cells labeled with MitoTracker-Red to characterise the mitochondrial aggregation phenotype in order to compare smORF loss-of-function phenotypes with the wild-type and diseased S2 cell mitochondrial populations.

The experiment was verified by the positive controls DIAP1, reaper and 14-3-3, which had a significant effect on the morphology of mitochondria as reflected by the high proportion of abnormal mitochondria in these samples, especially that of DIAP1, which

is known to induce apoptosis upon knock-down. Reaper is an inducer of apoptosis when expressed at high levels, but it also plays a role in regulating mitochondrial fusion and can be found localised to the mitochondria (Thomenius *et al.* 2011), and is therefore justified to shows an effect in this screen. Knocking down 14-3-3ζ has been shown to have an effect on phagocytosis (Ulvila *et al.* 2011), but this screen also reveals the majority of the treated cells exhibiting mitochondrial aggregation upon 14-3-3ζ RNAi, showing that when a master regulator of the cell is knocked down, the mitochondria will be affected. 14-3-3ζ knockdown in human cell lines has been shown to reduce cell growth and confer sensitivity of cells to apoptosis, whilst over-expression of 14-3-3ζ can define a risk of breast cancer as it promotes cell survival (Neal *et al.* 2009).

There is a high positive hit-rate for mitochondrial SEPs in this screen (12/23, 52%) as well as a high overall hit rate as 15 of the total 37 smORFs tested show a positive phenotype. Mitochondrial SEPs are highly enriched in the hit-list of the RNAi screen (80% of positive hit smORFs), which suggests that this screen directly reflects mitochondrial function. Therefore, it also emerged that the subcellular localisation of a smORF peptide is a precondition and hypothesis to be tested by the functional screen. That this screen more favourably detects phenotypes of mitochondrial SEPs, gives us an edge at determining which smORFs have a similar effect on mitochondria as proteins like DIAP1 and Drp1 which are directly related to apoptosis and mitochondrial function.

Some of the positive hits in this screen could be playing very important mitochondrial roles. *Drosophila* mutants of NADH complex 1 exhibit phenotypes that resemble a mitochondrial disease called Leigh syndrome. This disease is associated with decreased lifespan, seizures, heat intolerance, movement problems and progressive neurodegeneration (Burman *et al.* 2014). It would be very interesting to see whether the SEP from the smORF CG32230, which is a subunit peptide of NADH complex 1 and therefore also integral to metabolism, is involved in the prognosis of these symptoms. Another hit in this screen, which supports this hypothesis is Tim10, which is encoded by the dicistronic transcript consisting of the two mitochondrial SEPs, CG9878/CG42497. Tim10 is known to function as an inter-membrane translocase that transports molecules from the outer mitochondrial membrane to various processing enzymes on the inner

mitochondrial membrane (Schmidt *et al.* 2010). This helps me to conclude that this screen is reliably detecting the function of annotated smORF encoded peptides with a high level of reproducibility across repeats, despite the double-blind method of analysis and scoring the phenotype.

Knocking down Hemotin has quite a significant effect on mitochondrial morphology, even though there is substantial evidence to show that the Hemotin peptides are not generally found in the mitochondria. Hemotin contains a mitochondrial targeting motif (TargetP) in the peptide sequence, and Stannin has been found to localise to mitochondria in mammalian cells. This study shows that Hemotin may also be involved directly with the mitochondria, and this role has been previously unexplored by anyone.

Another interesting hit that emerges in this screen is CG14036, which may be important as it can be traced back to a study conducted on a genetic interactor of the *parkin* gene (Cox and Spralding 2009). The *Drosophila* gene 'clueless' encodes a CluA orthologue that is 53% identical to Clu in human. Mutations in *Drosophila clu* elicits mitochondrial dysfunction and clustering in a fashion similar to *parkin* mutants. *Clu* null mutants have a decreased life span, sterility and mitochondrial abnormality in flight muscles. *Drosophila* Clu interacts genetically with parkin and CG14036 is shown to be down-regulated in *Clu* heterozygous deficiency mutants (Cox and Spralding 2009). In my study, I have concluded that CG14036 deficiency also elicits a mitochondrial clustering response. As in the case with most of the hits in this screen, it may be that more mitochondrial SEPs are involved in the mitochondrial functional machinery, and by removing these smORFs may lead to mitochondrial phenotypes as severe as those found for canonical genes. Mutations in these smORFs would inevitably affect long term viability of the animal in which they are found, and could be involved in many pathways related to mitochondrial disease.

Preliminary results from new Poly-Ribo-Seq experiments in our lab show that several smORFs translated in embryos overlap with those being expressed in S2 cells. The interesting fact about these overlapping smORFs is that they are expressed throughout embryogenesis (Personal communication MAS Mumtaz). The smORFs that are unique to translation in embryos are temporally regulated in their expression, and analysing those will most likely reveal smORFs involved in development and differentiation of

the embryo. As for the smORFs expressed throughout embryogenesis, their overall functions are probably involved in processing energy, cell survival, and overall cellular homeostasis. This is probably what the functions of S2 cell smORFs are related to, as these cells are not hindered by the patterning boundaries of the embryo or signalled to from other tissues to die or change, and their main purpose is to survive. In this way, cell lines are almost like tumours, as given enough media and time, they will grow until nutrients run out or when stress is applied experimentally.

Though this assay was more time consuming than a generic cell viability assay conducted in a 96-well plate, it resulted in a large number of hits out of the small number of smORFs that were tested. This assay is robust and reliable at what could be considered a low N number for a typical cellular study which may employ automated imaging and high content microscopy to image hundreds of cells per repeat. In my screen, 50-100 cells were imaged per repeat and resulted in consistent and reproducible results. Therefore, this screen should be applied to the larger pool of translated smORFs in S2 cells because a potentially large proportion of these smORFs could be localised to the mitochondria. This screen could also theoretically be used to verify the effect of smORF over-expression on mitochondrial morphology in S2 cells using the tagging-transfection method. This may lead to the discovery of other smORFs like Reaper, which, upon over-expression in S2 cells, leads to mass cell death (Thomenius *et al.* 2011).

**The Alamar Blue assay does not detect smORF RNAi phenotypes**

The AB assay was conducted in order to observe whether knocking down of smORF peptide transcripts would result in a decrease in cell viability through measurement of cellular metabolism. Cell viability can be interpreted by comparing wild-type and mutant percentage reduction of AB reagent which is read by fluorescence signal derived from AB reagent incubated with the dsRNA-treated cells. AB is reduced by the Hydrogen transport molecules found in the mitochondrial matrix (NADH, FADH) which conduct a constant process that occurs in healthy cells. The severity of the smORF loss-of-function phenotype could then be assessed by comparing it to AB

reduction of samples treated with RNAi against known genes affecting cell viability, in this case DIAP1 and cdc2.

At this stage, although we lacked the information of mitochondrial subcellular localisation and distribution of SEPs from the tagging-transfection assay, we did have the Poly-Ribo-Seq results showing that the smORFs tested in the AB assay were translated in our lab S2 cell population. The assumption that was been exercised at this at this initial stage was that these smORFs are probably not involved in 'life or death' processes as this would have led to discovery and investigation of these smORFs already. More likely, these smORFs are involved in homeostatic cellular processes, such as intracellular transport, like Hemotin, or transport of ions across membranes, such as Sarcolamban as they were expressed at high levels in a cell culture system. In this vein, AB appeared to be a good assay for smORF functional read-out, as knocking down proteins which shuttle nutrients or small molecules, would surely have an impact on the overall cellular metabolic process. However, the results from the AB experiment were difficult to discern.

Knocking down DIAP1 by RNAi induces apoptosis (Wang *et al.* 1999; Steller 2008) and RNAi of cdc2 is known to decrease cell number (Björklund *et al.* 2006). Though knock-down of DIAP1 and cdc2 was verified by RT-PCR, the RNAi against these transcripts in the AB assay did not show a reproducible or significantly different change in AB reduction than the negative control and wild type samples. Though this may have been due to not choosing the correct positive controls to show an effect on the reduction of AB, and perhaps a better control would be one that effects the metabolic pathways directly. Cells treated with DIAP1 dsRNA, may have been respiring at wild-type rates despite the induction of apoptosis, but as DIAP1 is a commonly used positive control in cell viability screens, it should have technically shown some effect of a decrease in cell viability (Boutros *et al.* 2004). Since we observe knock-down of DIAP1, it is likely that cells in the DIAP1 RNAi sample are lowered in viability but the assay fails to pick up this change. In addition, two mitochondrial SEPs (CG32230 and CG7630) which show an effect in the mitochondrial morphology RNAi screen do not show an effect in the AB assay, even though CG32330 is annotated as a NADH complex 1 subunit. This dsRNA would target a direct metabolic function yet does not show an effect in the AB assay.

Even though the experiments were repeated with more dsRNA and longer incubation times to allow cells to show a significant effect of the knockdown of controls, the AB screen was not fruitful in detecting any of the smORFs tested as significantly functional. The two positive hits that come out in 2 or 3 repeats of the Pilot AB study do not have functions related to mitochondria, or cell death pathways according to their annotation, therefore I cannot conclude whether these are real hits or false positives in the screen. As we hoped to conduct the RNAi screen in a high-throughput fashion, it would not have been feasible to verify the knock-down of all the smORFs that were tested, therefore the AB screen was abandoned due to lack of success. However, I able to gain valuable knowledge from these experiments about the process of conducting an RNAi experiment and that the knock-down of controls is successful by this method in my hands.

# Chapter VI - Discussion & Conclusions

smORF encoded peptides (SEPs) are gradually being shown to be involved in various physiological and developmental processes. DNA and RNA sequencing data has revealed that there is the potential for translation of hundreds or even thousands of smORFs in almost all multicellular model organisms, as well as in lower eukaryotes like yeast. Though some smORF genes are now being annotated as protein coding genes in genomes, SEPs are generally difficult to detect due to the limitations of standard biochemical techniques, leaving them largely uncharacterised in terms of their broader functions.

The current definition of a smORF gene is characterised by the presence of one or more ORFs in it's transcript that is 2-99 codons in length, thereby encoding a peptide which is <100 amino acids (Basrai *et al.* 1997). Given that there are hundreds-of-thousands of smORFs present in most genomes that may potentially encode functional peptides (Basrai *et al.* 1997; Ladoukakis *et al.* 2011), it is essential to differentiate between smORFs that are translated and those that are not. Due to low rates of peptide detection, it is not yet well understood whether all of these smORF sequences are translated. However, examples exist in *Drosophila* of functional SEPs as small as 11 or 29 amino acids that play key roles in the development and physiology of the fruit fly (Galindo *et al.* 2007; Magny *et al.* 2013). The implementation of Poly-Ribo-Seq in Drosophila S2 cells allowed us to asses which smORFs are translated in S2 cell transcriptome (Aspden *et al.* 2014). This study showed that annotated smORFs are translated in a similar proportion as standard protein coding genes (~85%), therefore we ended up with a list of at least 228 SEPs for which we need to ascertain function. This number exceeds the scope of the single-gene genetic analysis and in-depth characterisation approach used by our laboratory in whole organism studies.

The hypothesis tested in this thesis is whether smORF functionality can be assessed at the cellular level, which would provide a new method by which to assess SEP functions at higher throughput than the single-gene studies used previously. The results reported here show that this is indeed possible and that by studying subcellular localisation and loss-of-function phenotypes in cell culture, we can start to identify the specific functions

of a larger number of SEPs. This study brings forward novel functional data on a significantly larger number of *Drosophila* SEPs than has ever been collectively studied before.

**The Hemotin SEP is involved in intracellular trafficking**

My investigation into smORF function at the cellular level began with *hemotin*, (Chapter III) which is a dicistronic smORF gene that encodes two peptides of 88aa (ORF1) and 59aa (ORF2) in size. Other members of our lab characterised the temporal expression of *hemotin* through semi quantitative RT-PCR, to show increased expression at the developmentally significant pupal stage and in adult flies. In embryos, *in situ* hybridisation showed that the pattern of expression of *hemotin* mRNA was similar to that of haemocyte markers, suggesting that Hemotin could play a role in haemocytes. In Haemocytes are actively involved in early *Drosophila* development in addition to their better characterised later functions in the immune system. Haemocytes are highly phagocytic and motile cells, which perform chemotaxis during wound healing and in response to infection (Wood and Jacinto 2007; Moreira *et al.* 2011). Since *hemotin* mRNA can be detected in embryonic haemocytes, they were used as the system in which to study the function of this SEP. These cells were an ideal system to study the function of Hemotin, as they can be assayed both *in vivo* and *ex vivo* .

In order to prove that *hemotin* is present at the protein level, I attempted to detect the Hemotin peptide using some of the different approaches discussed below. On the expectation that Hemotin is expressed and functioning in haemocytes, a *UAS-Hemotin-GFP* fusion protein plasmid was used to explore the subcellular localisation of Hemotin peptides with haemocyte-specific GAL4s to target the expression of this construct. This expression was readily detected by labelling with GFP specific antibodies, showing that the *hemotin* ORF1 is capable of producing a stable peptide product. I also attempted to detect endogenous expression of Hemotin in embryos and L3 and WPP haemocytes using a polyclonal antibody raised against Hemotin. Fluorescent antibody staining was used to verify endogenous Hemotin expression in embryos, and to test whether it can be detected in *ex vivo* haemocytes extracted from the animal at life stages where expression of the hemotin transcript has been detected by RT-PCR, and RNA-Seq by the

modENCODE project (L3 and WPP). Persisting signal from the antibody in the negative control samples (*Hemotin* null-mutant embryos and haemocytes) reinforced prior suspicions by our lab that antibodies raised against smORF peptides are generally inefficient. Direct antibody labelling has proven to be an experimental method that fails to detect endogenous smORF-encoded peptides. In the case of *tarsal-less,* the specific antibody also did not detect the endogenous peptide, but did manage to detect the much larger Tal-GST fusion protein. In the case of Hemotin, the Hemotin antibody staining did not colocalise to the signal from Hemotin-GFP expressed in haemocytes. Although this prevents us from knowing whether endogenous Hemotin protein is found in any other organs in the fly, the use of the tagged-construct provides proof of translation of Hemotin-ORF1 *(CG43194).* This also allowed us to move forward in characterising the function Hemotin performing in haemocytes through alternative techniques.

Using the Hemotin-GFP fusion protein, I carried out colocalisation analysis with antibodies and markers of ER, Golgi and transport vesicles (anti-KDEL), mitochondria (anti-GRP75) and acidic organelles such as Lysosomes and lysosome precursors (LysoTracker-RED). These experiments revealed that Hemotin-GFP colocalises strongly to KDEL and Lysotracker at both L3 and WPP stages. This suggests that Hemotin is performing a role in intracellular trafficking or endocytosis based on the knowledge that these are the main functions related to the organelles where Hemotin-GFP localises in haemocytes.

This discovery led to a series of phenotypic studies to confirm the role of Hemotin in these processes. In order to achieve this, I had to search for a phenotype related to these functions in haemocytes in which *hemotin* is removed from the flies using genetics. If this phenotype occurs specifically due to the lack of Hemotin being expressed, it can then be used to find the pathways or interacting partners of Hemotin function by comparing Hemotin mutant phenotypes to wild type animals or known mutants of that specific function. A *Hemotin*-deficiency mutant (*Df A4*) was created by E.G. Magny, who had observed that flies develop into morphologically wild-type looking adults under normal growth conditions. With no overt phenotype related to the *hemotin* mutation, more in-depth work was required to discern the function of this smORF.

The following work was conducted to see what would happen to the functioning of haemocytes in the absence of Hemotin expression. The target of these assays were the processes of intracellular trafficking or immunity, which could be related to the possible functions of a peptide localised to KDEL and Lysotracker-specific organelles in haemocytes. First, I conducted viability studies to see whether these flies have similar life-spans to wild type flies. *Df A4* flies have decreased viability at late adult stages and this effect is exacerbated upon infection with a large load of inert bacteria. Wild type flies survive this non-pathogenic bacterial immune challenge and live almost as long as untreated flies. *Df A4* homozygous flies begin to show signs of trouble much earlier than wild type flies. This result supports a role for Hemotin in the correct functioning of haemocytes in immunity.

The phagocytic/endocytic role of haemocytes is not just needed for immunity but comes into a developmental context for the clearance of apoptotic cell debris during embryonic and pupal stages (Wood and Jacinto 2007). These stages are characterised by the differentiation and re-modelling of body tissue as various organs are prepared for the next stage of life, therefore many cells undergo apoptosis at these times and leave behind cellular debris. Imaging the haemocytes in live embryos and pupae of *Df A4* mutants revealed that there are significantly larger vacuoles, or cellular inclusions, in *hemotin* deficient haemocytes. Live-imaging of pre-pupal haemocytes cultured *ex vivo* revealed that these cells contain significantly larger Lysotracker-labelled organelles than wild type pre-pupal haemocytes. Lysotracker labels late-endosomes and lysosomal compartments in the cell that have an acidic pH due to the acidifying enzymes responsible for breaking down phagocytised material in haemocytes (Watts 2012). In *hemotin* mutants, larger lysosomes persist in size, while wild type cells display a cyclic increase and decrease of lysosomal size in the same length of time. These data support a role for Hemotin in endocytosis and intracellular trafficking, along with the colocalisation of Hemotin-GFP to Lysotracker and KDEL markers.

*Ex vivo* characterisation of *hemotin* mutant haemocytes could help to provide an easily quantifiable phenotype for Hemotin. This study shows that *hemotin* mutant haemocytes display a largely non-motile population of cells at WPP stage, unlike wild type cells which are mostly motile at this stage. Previous characterisation of haemocytes showed

that they are active at this stage and display directional motility (Sampson *et al.* 2013). A non-motile profile has been observed in haemocytes that are mutant for cytoskeleton-regulators and phagocytic pathway genes. These cells cannot move because they struggle to recycle internalised membrane from phagosomes to the cell edge in order to extend filopodia and lamellipodia (Velichkova *et al.* 2010; Sampson and Williams 2012). The *hemotin* mutant cellular profile could only be partially rescued by expressing *UAS-hemotin-ORF1* and *UAS-hemotin-full length* constructs with haemocyte-specific GAL4s in the null mutant background. It was later found that this motility phenotype could also be attributed to the removal of another gene (*CG7691*) in the hemotin-null (*Df A4)* line (C.J. Sampson). As the non-motility of *hemotin* mutant haemocytes was not exclusively due to loss of Hemotin, it explained why only a partial rescue of the phenotype could be observed by expressing *hemotin* in a null background. Therefore we could not use this assay as a test for Hemotin function.

Based on the colocalisation studies, the large vacuoles in mutant embryo and pupal haemocytes, and the large lysosomes in *ex vivo* haemocytes, the focus remained on the role for Hemotin in intracellular trafficking,. Work done by J.I Pueyo in the laboratory using *ex vivo* cultured haemocytes showed that the enlarged vacuole phenotype is specific to Hemotin by rescuing expression of hemotin in a null background (A4 *Df)*. The decrease in viability after infection shows that there is a role for Hemotin in overall immunity of the fly, and that this is not attributed to CG7691 (Chapter III; Figure 3.7). I postulate that by removing Hemotin, we hamper intracellular trafficking, which in turn leads to the cell being overwhelmed and unable to digest pathogens that have been phagocytised. There is no obvious developmentally relevant morphological phenotype of Hemotin in the fly but studying it at a cellular level has led us to understand the function of this SEP. The role of Hemotin would be important in any cell type which is capable of performing endocytosis, especially in times of infection or when the cells are undergoing apoptosis, as intracellular trafficking mediates the removal of the by-products of these processes.

Running the peptide sequence of Hemotin through a homology-prediction software called PHYRE2 revealed a putative homologue for Hemotin called Stannin (Snn). Snn is an 88aa peptide similar to HemotinORF1. J.P. Couso has confirmed by sequence

alignment of this peptide across different phyla that the Hemotin peptide is conserved from *Drosophila* to mammals. Stannin was isolated from organotin toxin Trimethyltin (TMT) exposed cells in rat brain tissue (Toggas *et al.* 1992). *Snn-FLAG* fusion constructs expressed in rat cell lines showed that Snn colocalises to mitochondria and endoplasmic reticulum markers. Snn has also been isolated from the peroxisomal fraction of mammalian cells. Snn can be found (from high to low levels) in the spleen, brain, midbrain, kidney and lungs. It is widely expressed in the developing rat embryo then localises to the spleen and immune cells in the adult (Toggas *et al.* 1992; Dejneka *et al.* 1997). TMT exposure causes cell death specifically in cells that express this peptide and in this way, Snn confers toxin-sensitivity upon the tissue in which it is expressed. Snn expressing cells show increased vacuole formation in tubular epithelial cells. When lymphocytes lack Snn, this impairs immune function, causes splenic atrophy and results in altered cell cycle kinetics in mice (Billingsley *et al.* 2006). This indicates that there may be functional homology between Snn and Hemotin through vacuole formation and immunity.

*Stannin* expression is up-regulated by the cytokine TNF-alpha, which is expressed in human cell lines and tissue upon stress and inflammation (Horrevoets *et al.* 1999). The peptide sequence of Snn contains a conserved CXC metal-binding motif (cytokine receptor) as well as a putative 14-3-3ζ binding domain. 14-3-3ζ is a soluble, dimeric protein with known binding to the RXS+(pS/G**)**XP and RSESEE motif, which is found in the Snn peptide sequence. 14-3-3ζ co-immunoprecipitates with Snn-FLAG expressed in PC-12 rat cells, showing that these peptides are interacting physically (Davidson *et al.* 2005). Hemotin-GFP overlap to a pan-14-3-3 antibody could be observed in my experiments, supporting the homology between Hemotin and Snn. Full functional analysis has now characterised the interaction between 14-3-3ζ and Hemotin (Pueyo *et al.* 2016, in press). Hemotin represents yet another smORF peptide characterised in *Drosophila* that acts as a regulator of a specific process (endocytosis) by interaction with a larger molecule (14-3-3ζ) in a specific tissue (haemocytes), similar to Sarcolamban which interacts with SERCA in heart muscle (Magny *et al.* 2013). The study of Hemotin function shows that a lack of an overall body morphology phenotype

does not equal to a lack of smORF functionality at the cellular level, which was the case for Sarcolamban as well.

Understanding the functional role of a smORF by using fly genetics is currently dependent on the ability to detect the peptide in a particular tissue or developmental stage, such as haemocytes, and then characterise the loss-of-function phenotype in processes related to those clues, such as immunity or phagocytosis in this case. In addition to the challenge of defining smORF peptide translation, the wide range of functions (as seen by GO term analysis of translated smORFs) (Ladoukakis *et al.* 2011; Aspden *et al.* 2014) predicted for smORFs means that straightforward, high-throughput genetic analysis was not considered possible. However, by attempting smORF functional characterisation in cells and then using the information gathered to explore their function in the fly, can facilitate a more in-depth characterisation of SEP function on a wider scale.

**Subcellular localisation of Poly-Ribo-Seq smORF peptides reveals mito-SEPs**

The tagging-transfection assay (Chapter IV) corroborated the translation of the entire pool of smORFs tested (45 smORFs), which is 20% of the total high-confidence smORFs found translated by Poly-Ribo-Seq (228) in S2 cells. The tagging-transfection assay used heterologous gene expression in S2 cells to express FLAG-tagged smORF DNA under the actin 5c promoter. This method is considerably convenient to conduct in tissue culture, as it only requires the mixing of DNA and cells in the media. Creating a transgenic fly line with a fusion construct plasmid requires much more work, and three generations must pass before experiments can be conducted. Embryos must first be injected with the UAS construct, and then a stable fly line is created by mapping the insertion to a chromosome in order to further cross it to flies containing specific GAL4s to drive expression of the construct. Transient cell transfection bypasses all these steps by allowing transfection and assays to be conducted on a single population of cells in a plate. Given the challenge of generating efficient antibody against SEPs, alternative methods such as these have been used so far provide proof of translation and label the peptide (Galindo *et al.* 2007; Magny *et al.* 2013; Slavoff *et al.* 2012; Vanderperre *et al.* 2013; Slavoff 2014;).

The use of fluorescent imaging to detect the presence of FLAG-tagged peptide in the cells provided an opportunity to visualise the subcellular distribution and localisation of the SEPs. I observed four distinct categories of SEP distribution: 'Reticular', 'Cortical', 'Limited' and 'Other cytoplasmic'. This variety supports the robustness of the tagging-transfection assay to assess SEP localisation and shows that this method is probably not creating an artifactual bias to localisation. Categorising tagged-SEPs into subcellular distribution classes allows a quick preliminary assessment of potential localisation of SEPs, and could be used as an initial tool for assessing smORF function in a semi-high-throughput manner. The largest group of subcellular distribution are the 'Reticular' SEPs, which localise completely to the mitochondria (Colocalisation to MitoTracker-Red, Chapter IV). 'Other cytoplasmic' SEPs also show a medium level of colocalisation to the mitochondria, and have Mander's coefficients similar to Sarcolamban-ORFA (Scl-A) in this experiment. The 'Other cytoplasmic' SEP's have similar subcellular distributions to Scl-A, which suggests that they could be localised to the ER. The perceived overlap to MitoTracker could be due to the close-knit juxtaposition of the ER and mitochondria in S2 cells. 'Cortical' SEPs could be observed to localise near the plasma membrane edge of the cell, where filamentous actin forms a scaffolding around the cytoplasm. 'Cortical' SEPs show a partial overlap to F-actin at the edge of the cytoplasm, but the average pixel-to-pixel overlap of these SEPs to Phalloidin-Rhodamine was too low to conclude that they are predominantly found at the F-actin periphery. This shows that not all of the peptide is localising to the cell edge, but could also be involved in another cytoskeletal-associated or intracellular capacity. For example, in this assay both tagged ORFs of Hemotin appear to be Cortical SEPs, but I and others have shown that the role of Hemotin is in intracellular organelles. This raises the question of whether the high levels of expression due to the the actin 5c promoter may be distorting the localisation of 'Cortical' SEP's by spreading their expression beyond their actual sites of wild-type function, which may also mean that the amount of expression of the other tagged-SEPs may be similar to this scenario. 'Limited' SEPs show colocalisation to F-actin at similar levels to 'Cortical' SEPs, though their observed subcellular localisation is very different to the eye. Studies on the 'Limited' SEP's should be followed up by quantifying co-localisation with internalised membrane

markers such as FM464, to show whether or not this is true. The partial overlap of 'Limited' smORFs to F-actin could be because these SEPs are in intracellular vesicles and they may have a role in digestion.

Amino acid length distributions of the different subcellular distribution classes showed that 'Reticular' and 'Cortical' SEPs tend to be near the upper-limit of SEP size (>50 amino acids), whilst 'Other Cytoplasmic' smORFs have a wide distribution of amino acid lengths. 'Limited' smORFs are distinctly smaller in size than all the other distribution types and most of the 'Limited' SEP's are derived from smORF sequences found within ncRNAs. It would be interesting to test whether the amino acid length of an SEP can predict the subcellular distribution of SEPs, using a size-selected sample of smORFs from the translated pool. The amino acid sizes of Limited SEPs falls at the range of what are called 'dwarf smORFs' which are approximately 20aa. Dwarf-smORFs are predominantly found in 5'UTRs and ncr_ORFs. They are not well conserved and there is generally no bioinformatic indicators of peptide function (Aspden *et al.* 2014).

As the Reticular smORFs were over-represented in the tagged-smORF pool, they became the focus of further functional characterisation. The overlap of Reticular SEPs to the mitochondrial marker MitoTracker-Red was even higher than the overlap between the positive control (*UAS-mito-GFP* expressed in haemocytes) and MitoTracker dye. If the results of the tagging-transfection assay are extrapolated to the larger pool, it appears that a very large proportion of S2 cell SEPs have mitochondrial localisation (44%). This is more than the proportion of canonical mitochondrial protein-coding genes, which are placed around 10-15% of total protein coding genes (Neupert and Herrmann 2007). GO term molecular annotations of the Poly-Ribo-Seq smORFs show an enrichment of catalytic activity (oxidoreductase) and transmembrane transporters, though the number of SEPs with annotation is low (Aspden *et al.* 2014). It is possible that Poly-Ribo-Seq smORFs actually contain very high numbers of mitochondrial SEPs, but TargetP (Emanuelsson *et al.* 2007) only accurately predicted these half of the time when compared to experimental verification by the tagging-transfection assay.

**Subcellular localisation to functional assessment**

Understanding why so many smORFs are mitochondrial can place us one step ahead in characterising the important cellular function of SEPs, but in order to do that, we need confirmation of which smORFs actually elicit a mitochondrial stress response when they are removed from the cell. Therefore, it would be beneficial to know whether they localise to the inner or outer mitochondrial membranes. Even though both the inner and outer mitochondrial membranes are composed of a phospholipid bilayer, they are unique in protein composition and permeability, which relates to their distinct functions. Conventional laser confocal microscopy does not provide enough resolution to determine whether these peptides are embedded in the membranes of these organelles, but it could be very likely that some SEPs are localised there. The outer mitochondrial membrane (OMM) has a 50/50 ratio of proteins to phospholipid, whilst the inner membrane (IMM) is composed mainly of proteins (80%) (Krauss 2001). The OMM is permeable to ions and peptides of about 50aa and smaller (Lodish 2008; Park *et al.* 2010). The median length of our Reticular SEPs is higher than this 50aa limit (66.79±22.19aa), showing that most of the Reticular SEPs cannot just passively diffuse into the mitochondria from the cytosol and must be localised there for a specific purpose. Required molecules that are larger than 50aa are actively transported to the inner membrane by Voltage Dependent Anion-selective Channels (VDAC) porins of the OMM (Young *et al.* 2007). Porins are conserved from bacterial to eukaryotic cells and studying them has revealed a common structurally property of VDAC porins; they consist of B-barrel structures and generally lack membrane-spanning alpha helices. Porins also allow the exchange of small molecules and ions like NADH and ATP across the OMM (Young *et al.* 2007; Park *et al.* 2010).

The IMM has very low permeability for ions and molecules, allowing for the discrete separation of the cellular cytoplasm and the mitochondrial matrix, where electron and proton transport occur and a membrane potential is required for cellular metabolism. The IMM contains the OXPHOS enzymes and ion transporters to only allow specific molecules to cross the IMM, such as the nuclear-encoded phosphate-OH2 exchanger, the adenine nucleotide translocase (which exchanges ADP for ATP) as well as fatty acid transporters and aspartate/glutamate shuttles (Krauss 2001). Our tagged-smORF pool

contains three annotated SEPs which have GO terms that place them on the inner membrane: Tim10 peptides CG42497 (65aa) and CG9878 (92aa) are part of an inner membrane translocase complex and CG32230 (83aa) is a NADH dehydrogenase subunit. Another interesting common feature of these SEP's is that they contain transmembrane alpha-helices and it could be possible that they play a role similar to the 79aa MTGM (mitochondrial targeting GxxxG motif) SEP recently characterised in mammals and conserved throughout eukaryotes. MTGM is a mitochondrial inner membrane protein, and contains a transmembrane alpha-helical domain (Zhao *et al.* 2009). The expression of this peptide increased reactive oxygen species (ROS) levels and caused DNA damage. MTGM depletion results in elongated mitochondria, decreased cell proliferation and resistance to apoptosis upon staurosporine exposure. The expression of this SEP confers a type of sensitivity to stress factors in the mitochondria, and regulates mitochondrial dynamics for optimal growth of both human brain tumour cell lines and tumour tissues (Zhao *et al.* 2009). The 63aa mitochondrial-located SEP, Boymaw also seems to perform a similar role in cells, where the SEP translated from ORF3 is translocated to the mitochondria after increased expression during stress (Ji *et al.* 2015). In the mitochondria, Boymaw inhibits oxidoreductase activity in order to promote cell survival. The small size of this peptide allows it to be quickly degraded after permissive levels of stress are achieved. Few SEP's have been previously linked to be directly impacting the function of mitochondria in *Drosophila*, but the precedence in mammalian cells makes the study of more mito-SEPs very valuable.

Mitochondria are very complex organelles and there are still many processes that occur on or inside the mitochondria that are not yet completely understood. Besides their primary role in energy production, mitochondria can serve as reaction centres for signal transduction of cell survival, death and proliferation (Tait and Green 2012). This makes sense as they are constantly monitoring the external and internal energy requirements of the cell. Mitochondria cannot be created *de novo* and each cell in an organism inherits mitochondria from predecessor cells, which themselves have been donated by the maternal parent (Jornayvaz and Shulman 2010). This has implications related to longevity and ageing as well as stress and infection, and older mitochondria are known

to cause a decrease in viability (Finkel 2015). Mitochondria dynamically respond to stress, and mitochondrial stress can be elicited by a variety of stimuli, such as toxification by paraquat (Ivatt *et al.* 2014), chemotherapy drugs, UV irradiation, Ca2+ misregulation (Frieden *et al.* 2004), starvation (Rambold *et al.* 2011), inhibition of cytoskeletal dynamics and a variety of genetic factors (Park *et al.* 2010; Westermann 2010; Westrate *et al.* 2014; Zemirli *et al.* 2014; Norris *et al.* 2015). Tubular elongation and clumping protects the mitochondria from autophagy and preserves mtDNA while still allowing respiration to occur in times of duress (Rambold *et al.* 2011; Gomes *et al.* 2011). When the stress is removed, the system should potentially return to normal. However, when the stress is prolonged, as it would be in the case of a genetic disease, this translocation response of mitochondria eventually leads to increased generation of oxygen radicals from oxidative stress (Redpath *et al.* 2013) leading to mtDNA damage. Eventually, this results in decreased ATP production as enzyme subunits encoded by mtDNA would not be efficiently translated. ATP is required for the continued function of all the other organelles in the cell. In neurodegenerative diseases, this could potentially lead to the accumulation of mis-folded proteins due to lack of available energy to remove them, which result in the plaques that are observed in neurons (Chen and Chan 2009). These plaques are the cause of synaptic loss in patients suffering from neurodegenerative disease, and mitochondrial dynamics can be related to several neurodegenerative diseases (Chen and Chan 2009; Su *et al.* 2010; Itoh *et al.* 2013). No matter what the organ, if imposed factors lead to mitochondrial stress in a certain tissue, it would eventually cause damage to that organ through the effect on mitochondria. When enough damage has been done to mitochondria and mtDNA to push the cells over the edge, mitochondrial fission occurs, leading to the release of apoptotic factors, which signal the transcription of cell death factors such as reaper, and leads to cell death(Suen *et al.* 2008; Westermann *et al.* 2010; Westrate 2014).

The discovery of several mito-SEPs made us question which specific characteristics of these peptides may cause them to localise to this organelle. As so many mitochondrial functions occur on membranes, it could be possible that SEPs are found in the mitochondrial membrane, acting as regulators of larger proteins, such as Sarcolamban or Hemotin. Approximately 30% of the total pool of tagged-smORFs contain high

confidence TMHMM alpha-helix predictions and overall, 32.3% of the Poly-Ribo-Seq translated smORFs have one or more predicted transmembrane alpha-helices, similar to the proportion of TMHMM in the total annotated FlyBase smORFs (32.2%). CG32230, CG7630, CG32276, CG34242 and the dicistronic CG15386 and CG9878 (Tim10) are mito-SEPs with TMHMM motifs. However, there are 2 examples of annotated mito-SEPs that localise to the mitochondria, but do not contain a predicted alpha-helix, namely CG14482 (57aa), a cytochrome b-c1 complex subunit which would potentially be found on the IMM and CG44242 (70aa), a mitochondrial ribosomal subunit, which would reside in the mitochondrial matrix. In addition, there are mito-SEPs with no TMHMM or functional prediction: CG33199 (79aa), CG33714 (90aa), CR43250 (73aa), CR43144 (91aa), CG14036 (93aa), CG14014 (68aa), CR32582 (52aa), and the dicistronic CG15386 (77aa) and CG42371 (87aa). Though subcellular localisation does indicate where these SEPs are found, there are very few other indicators of what function these latter mito-SEPs are involved in. Their sizes show that they cannot passively diffuse across the OMM, so they are probably playing a specific role in the mitochondria, but this needs to be verified by further functional testing.

**Mitochondrial morphology RNAi screen reveals Functional mito-SEPs**

The results from the tagging-transfection assay allow us to focus functional screening on specific cellular processes in which they may be involved. This led to the targeted testing of the function of these S2 cell Poly-Ribo-Seq translated smORFs in the mitochondria of S2 cells. The dynamic nature of mitochondria can be used as a reflection of the internal state of cellular health. Since the mitochondrial aggregation phenotype could be observed in our S2 cells, and so many of the tagged SEPs localised to the mitochondria I decided to design a screen based upon the mitochondrial stress response (Chapter V). Oxidative stress has been shown to lead to mitochondrial perinuclear clumping in cells (Redpath *et al.* 2013). The RNA-interference method degrades mRNA transcripts, leading to an indirect decrease in the peptide product of a target gene. This method became viable for the testing of smORF function after obtaining proof of translation of smORFs through Poly-Ribo-Seq (Aspden *et al.* 2014) and the tagging-transfection assay described in this thesis. Using the pool of tagged-

smORFs as a starting point, several dsRNAs were designed to knockdown expression of these smORFs.

This key phenotype exhibited by mitochondria under stress, is generally referred to as mitochondrial hyper-fusion or aggregation (Tondera *et al.* 2009). By screening for mitochondrial aggregation phenotypes after smORF RNAi, I hoped to see the effects of smORF knockdown on the organelle where so many are localised to and may be functioning in. The mitochondrial aggregation phenotype could be observed in our S2 cells at a basal level in about 12–20% of the cells that were imaged from wild type samples. The induction of a non-specific RNAi response by targeting the BlueScript plasmid resulted in mitochondrial phenotypes similar to dsRNA-untransfected cells. Positive controls of cell death, such as DIAP1 and reaper (inhibits mitochondrial fragmentation; 68aa SEP (Thomenius *et al.* 2011)), caused levels of abnormal mitochondria to rise to a significantly higher level when compared to negative-control treated cells. Several major regulators of mitochondrial dynamics have been identified and the key players in these processes have been studied in high detail (Karbowski and Youle 2003). Knocking down one of these major regulators, Dynamic-Related-Protein (Drp-1), is known to induce a mitochondrial hyper-fusion response (Qian *et al.* 2012). This was also the case in my screen, as the dsRNA designed against Drp1 raised the level of aggregated mitochondria to more than 60% of the imaged cells in these samples, offering a reliable positive control in the screen. RNAi against 14-3-3zeta, a major regulator of the cell cycle and intracellular trafficking resulted in similar phenotypes to dsRNA-Drp1. This duplex is targeted to a region common to all the isoforms of 14-3-3 which would also be very detrimental to the cell.

Now that an upper and lower limit of mitochondrial aggregation phenotypes had been established, the results from knocking down smORFs could be interpreted. Of the 37 smORFs tested in this screen, 15 showed a significant mitochondrial phenotype when compared to dsRNA-BlueScript treated cells. The effect on mitochondria may be due to knocking down SEPs that contain a TMHMM prediction, which could compromise membrane integrity of the mitochondria as opposed to having a direct effect on mitochondrial function. However, the opposite is true, of the positive hits in the screen, 43.47% contain TMHMM predictions; while 81.81% of the negatives in the screen

contain TMHMM predictions. We can clearly observe that there is an enrichment of mito-SEP hits in the positive hits of the screen as 12 of the 15 positive hits have some evidence of mitochondrial localisation. Together, this information supports the idea that this screen is not biased towards smORF peptides which contain TM-alpha helical structures. Taken along with the fact that 80% of the hits are mito-SEPs, we can conclude that this screen detects mitochondrial morphological phenotypes as a direct response of removing mitochondrial-associated SEPs. This result is supported by examining the known data for these 'Functional mito-SEPs'. ORF2 of the CG42497/ CG9878 dicistronic smORF transcript is annotated as Tim10 and the Tim9-Tim10 complex is responsible for the transfer of peptides through the mitochondrial inter-membrane space (Schmidt *et al.* 2010). Knocking down this transcript must have a direct effect on an important mitochondrial peptide, and the strong phenotype of RNAi against Tim10 is an internal proof of concept for this screen. Another positive control for a functional mitochondrial SEP is the RNAi against CG32230, which is homologous to NADH complex 1 subunit peptide, and therefore playing a role in the electron transport chain on the IMM. CG32330 also shows a phenotype in this screen.

All four dicistronic smORFs tested, of which three are Reticular SEPs directly observed to localise to the mitochondria, showed a phenotype in this screen. Surprisingly, RNAi against Hemotin (Chapter III) shows a strong phenotype in this assay. Although I did not observe Hemotin-ORF1 localising to mitochondrial markers (GRP75, MitoTracker), ORF2 of Hemotin (CG43210-59aa) contains a predicted mTP sequence (ORF1 (CG43194-88aa) has a predicted SP motif). Since the the entire hemotin transcript would be knocked down by the dsRNA, this result highlights that there could be a potential role for Hemotin-ORF2 in mitochondria. This is supported from the subcellular distribution of FLAG-tagged Hemotin-ORF2, where we could observe some punctate expression in the cytoplasm as well as a Cortical distribution near the F-actin periphery. Overlap to MitoTracker-Red is higher for Hemotin-ORF2 (Mander's Coefficient = 0.37±0.14) than Hemotin-ORF1 (0.07±0.13). The other two novel dicistronic smORFs (CG32736/CG42308 and CG15386/CG42371) have no annotated function and could be very interesting for further study in the fly. Both of these dicistronic mitochondrial smORF transcripts are expressed ubiquitously in embryos,

larvae and adult flies and organs at quite high levels (modENCODE). These mito-SEPS could potentially be very important for cellular homeostasis based upon their localisation, phenotype and expression data. In addition, the monocistronic smORFs CG7630, CG33199, CG14036, CG14014, CR43144, CR43250 and CG33714 are positive hits in the mitochondrial morphology screen, with Reticular distribution and mitochondrial localisation, but their mitochondrial functions have never been verified in *Drosophila* or higher animals. The fact that two of these are annotated ncRNA transcripts containing translated smORFs also makes them good candidates for further study. The mitochondrial morphology RNAi screen detects mitochondrial function of 32% of the smORFs that were tested in the screen. Considering that the results are quite consistent between repeats for both test and control samples, it seems very likely that these smORFs are inducing a genuine stress response from the mitochondria, due to a direct functional link to this organelle.

Characterising loss-of-function phenotypes using a sensitive read-out is important, especially in the case of smORFs, as I discovered in my initial attempt at an RNAi screen with the fluorometric reagent called Alamar Blue (AB). AB measures the reducing environment of cells and would theoretically allow us to monitor the effects of knocking down a smORF by the perturbations that would occur in the metabolic state of the dsRNA-treated cells after inducing RNAi. The candidate smORFs for this screen were those found in the Peptide Atlas database (Desiere *et al.* 2006), but lacked information for the subcellular localisation of these peptides. Though two positive controls of cell viability (DIAP1 and cdc2) were chosen, neither gave a significant result in the many repeats of the AB assay. The results from smORF dsRNA samples were also inconsistent and it was concluded that our AB assay was not reliable. The two smORF hits (CG11985 and CG17059) that emerged as statistically different to wild type samples, across three AB experiments, were not predicted to have any function that is related to metabolic functions or the mitochondria. Furthermore, CG7630 and CG33714 which are now known mito-SEPs, do not show a phenotype in this screen. These results imply that wide-spread metabolic failure seems to be the last response to cellular stress, and may only occur as a consequence of a very strong phenotypic response from RNAi. This can be related to the result observed in DIAP1-dsRNA treated cells, which should

be undergoing apoptosis due to the removal of DIAP1. However, these cells exhibit wild type levels of reduction of the Alamar Blue reagent in the assay despite verified knock-down of DIAP1 by RT-PCR of DIAP1-dsRNA treated S2 cells. However, the mitochondrial morphology screen also produces some false negatives (mitochondrial-localised SEPs that do not show a phenotype), this implies that the RNAi technique itself may require further optimisation. It could also be that some target genes require a larger amount of or a longer incubation time with the dsRNA to effectively knock down the targeted smORF gene product. Ideally, we would verify this for individual smORFs by quantitative RT-PCR.

Mitochondrial morphological changes are involved in phenotypes observed in Alzheimer's Disease, Huntington's disease, Parkinson's disease, cerebellar degeneration, and peripheral nerve dysfunction (Charcot-Marie-Tooth disease) (Chen and Chan 2009; Grünewald *et al.* 2010; Itoh *et al.* 2013). Mitochondria play a significant role in the neuropathy of all the major neurodegenerative diseases, for which many of the genetic-risk loci currently have not been identified. All of these diseases have been linked to the induction of oxidative stress in the cells that are affected (Lin 2006; Federico *et al.* 2012). Could it be that old-age related diseases can be caused by genomic elements as small as smORFs? If true, this would change the way that we look for markers of these diseases, and could present novel therapeutic targets to alleviate symptoms. The smORF peptides that are positive hits in the RNAi screen may be performing functions of cellular homeostasis in wild type animals, but when they are removed, they cause mitochondria to translocate in a similar fashion as knock-down of very important proteins such as DIAP1, reaper, Drp1 and 14-3-3.

**Cellular characterisation of smORF function: a pipeline in progress**

The results in this thesis, from the initial characterisation of Hemotin to the discovery of 20 mitochondrial-SEPs in S2 cells, and the functional verification of 12 of those mito-SEPs by RNAi, highlights the advantages of studying smORF function at the cellular level prior to carrying out a full scale organismal study (Figure 6.1). These studies provide preliminary results and hypotheses to be corroborated by further study in the whole fly. Though these results are based on a small subset of the potential peptide

encoding smORFs in the *Drosophila* genome, they represent almost 20 percent of the total smORFs that are translated in S2 cells, and thus can be considered a representative sample. The Hemotin study showed that the use of epitope-tagged smORFs in haemocytes can help unravel its function through imaging. Though this thesis focused on the Reticular mito-SEPs, the staining technique allows labelling of different organelles, and using markers for the ER, Golgi and lysosomes. Therefore we could potentially confirm whether 'Other Cytoplasmic' SEPs are found in the ER, or whether 'Limited' SEPs are localised to intracellular vesicles and whether or not there is a pattern to these SEP containing a Signal-Peptide (SP) motif, as is the case for Hemotin-ORF1. There are published methods by which to assay whether or not a peptide is involved in the secretory pathway or the organisation of Golgi bodies using RNAi (Bard *et al.* 2006; Kondylis *et al.* 2011).

In order to truly expand our knowledge of the proteome on a higher-throughput basis, I believe that the cellular characterisation of smORF peptides is necessary and valuable. The developmental biology field has been saturated by the identification of novel developmental genes in the past two or three decades. However, in recent years, there is a trend towards the detection and characterisation of smaller genomic elements, which play subtle roles in the way that developed organisms and cells are being impacted by their environment. Micro-RNAs have been shown to regulate adaptive responses by affecting the expression and repression of important genes according to the environment (Diawara *et al.* 2014; Ma *et al.* 2011). Small peptides encoded by smORF genes in 5'UTR's (uORFS) play regulatory roles in the translation of downstream proteins (Child *et al.* 1999; Zhang 2005). The peptides encoded by *sarcolamban* and *Hemotin* also seem to be involved in regulatory roles with other proteins, acting upon specific organelles at the cell or tissue level. This thesis therefore contributes significantly to expanding the small pool of functional SEPs. These results can also aid in the definition of novel sequence motifs to aid the identification of more smORFs or SEPs.

Arising from my studies, I propose that the cellular characterisation of smORFs would use a pipeline of semi high-throughput techniques: 1) evidence of translation by Poly-Ribo-Seq; 2) confirmation of translation and study of subcellular localisation by the tagging-transfection assay; and 3) based upon subcellular localisation, assessment of

organelle-related loss-of-function phenotypes by RNAi screening. Using this pipeline can help to gather functional information about smORFs on a much wider scale than before. This data could result in smORFs and their SEPs being considered by others to unravel their animal physiological function in more detail.

| smORF | Length (aa) | Tm alpha helix | Poly Ribo Seq Tranlated | Peptidomic Evidence | FLAG-tag translation | TargetP | Distribution | Phalloidin-Rhodamine Overlap | MitoTracker-Red Overlap | Phenotype |
|---|---|---|---|---|---|---|---|---|---|---|
| CG9878 | 92 | 1 | Yes | No | Yes | Signal | Other Cytoplasmic | Low | High | Yes |
| CG42308 | 61 | 1 | Yes | No | Yes | Signal | Reticular | Low | High | Yes |
| CG7630 | 90 | 1 | Yes | Pep. Atlas | Yes | Mito | Reticular | Low | High | Yes |
| CG33199 | 79 | 0 | Yes | Pep. Atlas | Yes | Mito | Reticular | Low | High | Yes |
| CG33714 | 90 | 0 | Yes | Pep. Atlas | Yes | Mito | Reticular | Low | High | Yes |
| CG42371 | 87 | 0 | Yes | No | Yes | Mito | Reticular | Low | High | Yes |
| CG32736 | 79 | 1 | Yes | No | Yes | Mito | Reticular | Low | High | Yes |
| CG32230 | 83 | 1 | Yes | Pep. Atlas | Yes | - | Reticular | Low | High | Yes |
| CG14104 | 68 | 0 | Yes | No | Yes | - | Other Cytoplasmic | Low | High | Yes |
| CG14036 | 93 | 0 | Yes | Pep. Atlas | Yes | - | Cortical | Medium | High | Yes |
| CR43144 | 91 | 0 | Yes | No | Yes | - | Reticular | Low | High | Yes |
| CR43250 | 73 | 0 | Yes | No | Yes | - | Reticular | Low | High | Yes |
| CG42497 | 65 | 0 | Yes | No | Yes | - | Reticular | Low | High | Yes |
| CG43194 | 88 | 1 | Yes | Pep. Atlas | Yes | Signal | Cortical | Medium | Low | Yes |
| CG12384 | 96 | 0 | Yes | Pep. Atlas | Yes | - | Cortical | Medium | Low | Yes |
| CG43210 | 59 | 1 | Yes | Pep. Atlas | Yes | Mito | Cortical | Medium | Medium | Yes |
| CG34200 | 52 | 0 | Yes | Pep. Atlas | Yes | - | Limited | Low | Medium | Yes |
| CG15386 | 77 | 0 | Yes | No | Yes | Signal | Other Cytoplasmic | Low | Not tested | Yes |
| CG11686 | 70 | 1 | Yes | Pep. Atlas | Not Tagged | - | Unknown | Unknown | Unknown | Yes |
| CG34242 | 58 | 1 | Yes | No | Yes | Signal | Reticular | Low | High | No |
| CG44242 | 70 | 0 | Yes | Couso Lab | Yes | Mito | Reticular | Low | High | No |
| CG14482 | 57 | 0 | Yes | Pep. Atlas | Yes | Mito | Reticular | Low | High | No |
| CG32276 | 64 | 1 | Yes | No | Yes | Mito | Other Cytoplasmic | Low | High | No |
| CR32582 | 52 | 0 | Yes | No | Yes | Mito | Reticular | Low | High | No |
| CG17278 | 80 | 0 | Yes | No | Yes | - | Reticular | Low | High | No |
| CR30055 | 2-53, 4-17, 5-56 | 0 | Yes | No | Yes | - | Reticular + Limited | Low (ORF1) | High (ORF1) | No |
| CG33155 | 60 | 1 | Yes | Pep. Atlas | Yes | - | Limited | Low | Low | No |
| CG33170 | 71 | 0 | Yes | No | Yes | Signal | Reticular | Low | Medium | No |
| CG33774 | 40 | 1 | Yes | Pep. Atlas | Yes | Signal | Other Cytoplasmic | Low | Medium | No |
| CG32267 | 49 | 1 | Yes | Pep. Atlas | Yes | Mito | Other Cytoplasmic | Low | Not tested | No |
| CG6770 | 69 | 0 | Yes | Pep. Atlas | Yes | - | Cortical | Medium | Not tested | No |
| pncr009:3L | 1-21, 2-20 3 - 33 | 0 | Yes | No | Yes | - | Other Cytoplasmic + Limited | Low | Not tested | No |
| Rox2 | 1 - 38, 2-17 | 0 | Yes | No | Yes | - | Limited | - | Not tested | No |
| CG13784 | 67 | 2 | Yes | Pep. Atlas | Not Tagged | Signal | Unknown | Unknown | Unknown | No |
| CG17680 | 97 | 1 | Yes | Pep. Atlas | Not Tagged | Mito | Unknown | Unknown | Unknown | No |
| CG4101 | 87 | 1 | Yes | Pep. Atlas | Not Tagged | - | Unknown | Unknown | Unknown | No |
| CG18343 | 99 | 2 | Yes | Pep. Atlas | Not Tagged | - | Unknown | Unknown | Unknown | No |
| CG31360 | 98 | 1 | Yes | Pep. Atlas | Not Tagged | - | Unknown | Unknown | Unknown | No |
| CG9669 | 78 | 2 | Yes | Pep. Atlas | Not Tagged | - | Unknown | Unknown | Unknown | No |
| CG32248 | 82 | 1 | Yes | Pep. Atlas | Not Tagged | - | Unknown | Unknown | Unknown | No |
| CG15456 | 95 | 0 | Yes | Pep. Atlas | Yes | - | Cortical | Medium | Medium | - |
| Scl-A | 28 | 1 | No | Yes | Yes | - | Other Cytoplasmic | not tested | Medium | - |
| CG11985 | 85 | 0 | Yes | Pep. Atlas | Not Tagged | - | Unknown | Unknown | Unknown | Yes - AB |
| CG17059 | 57 | 0 | Yes | Pep. Atlas | Not Tagged | - | Unknown | Unknown | Unknown | Yes - AB |
| CG11284 | 16 | NA | Yes | No | Yes | - | Cortical | Medium | Not tested | - |
| CrebA | 16 | NA | Yes | No | Yes | - | Reticular | - | Not tested | - |
| FBtr0072084 | 14 | NA | Yes | No | Yes | - | Reticular | Low | Not tested | - |
| FBtr0072210 | 13 | NA | Yes | No | Yes | - | Other Cytoplasmic | Low | Not tested | - |
| FBtr0081720 | 11 | NA | Yes | No | Yes | - | Limited | Low | Not tested | - |
| tal-B | 49 | 0 | No | No | Yes | - | Other Cytoplasmic | Low | Not tested | No |
| tal full length | 49 | 0 | No | No | No | - | N/A | not tested | Not tested | No |

**Figure 6.1 Summary of Cellular characterisation of smORF functions in *Drosophila melanogaster***

# References

**Abdelwahid, E., Yokokura, T., Krieser, R.J., Balasundaram, S., Fowle, W.H. & White, K.**, 2007, Mitochondrial disruption in Drosophila apoptosis, *Developmental cell*, 12(5), 793–806.

**Aboobaker, A.A., Tomancak, P., Patel, N., Rubin, G.M. & Lai, E.C.,** 2005, Drosophila microRNAs exhibit diverse spatial expression patterns during embryonic development, *Proceedings of the National Academy of Sciences of the United States of America*, 102(50), 18017–18022.

**Akbar, M.A., Tracy, C., Kahr, W.H. & Krämer, H.**, 2011, The full-of-bacteria gene is required for phagosome maturation during immune defense in Drosophila, *The Journal of Cell Biology*, 192(3), 383–390.

**Anderson, D.M., Anderson, K.M., Chang, C.L., Makarewich, C.A., Nelson, B.R., McAnally, J.R., Kasaragod, P., Shelton, J.M., Liou, J., Bassel-Duby, R. & Olson, E.N.,** 2015, A micropeptide encoded by a putative long noncoding RNA regulates muscle performance, *Cell*, 160(4), 595–606.

**Andrews, S.J. & Rothnagel, J.A.**, 2014, Emerging evidence for functional peptides encoded by short open reading frames, *Nature Reviews Genetics*, 15(3),193–204.

**Arava, Y., Wang, Y., Storey, J.D., Liu, C.L., Brown, P.O. & Herschlag, D**., 2003, Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae, *Proceedings of the National Academy of Sciences of the United States of America*, 100(7), 3889–3894.

**Aspden, J.L., Eyre-Walker, Y.C., Phillips, R.J., Amin, U., Mumtaz, M.A.S., Brocard, M. & Couso, J.P.**, 2014, Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq, *eLife*, 3:e03528.

**Bard, F., Casano, L., Mallabiabarrena, A., Wallace, E., Saito, K., Kitayama, H., Guizzunti, G., Hu, Y., Wendler, F. & DasGupta, R.**, 2006, Functional genomics reveals genes involved in protein secretion and Golgi organization, *Nature*, 439(7076), 604–607.

**Basrai, M.A., Hieter, P. & Boeke, J.D.**, 1997, Small open reading frames: beautiful needles in the haystack, *Genome research*, 7(8), 768–771.

**Bate, M. & Martinez Arias, A.**, eds., 2009, *The development of Drosophila melanogaster,* Vol. 1, Cold Spring Harbor Laboratory Press, Woodbury, N.Y.

**Baum, B. & Cherbas, L.**, 2008, Drosophila cell lines as model systems and as an experimental tool, *Methods in Molecular Biology*, Humana Press, 420, 391–424.

**Bazzini, A.A., Johnstone, T.G., Christiano, R., Mackowiak, S.D., Obermayer, B., Fleming, E.S., Vejnar, C.E., Lee, M.T., Rajewsky, N., Walther, T.C. & Giraldez, A.J.**, 2014, Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation, *The EMBO Journal*, 33(9), 981–993.

**Bernstein, E., Caudy, A.A., Hammond, S.M. & Hannon, G.J.**, 2001, Role for a bidentate ribonuclease in the initiation step of RNA interference, *Nature*, 409(6818), 363–366.

**Bettencourt-Dias, M., Giet, R., Sinka, R., Mazumdar, A., Lock, W.G., Balloux, F., Zafiropoulos, P.J., Yamaguchi, S., Winter, S. & Carthew, R.W.**, 2004, Genome-wide survey of protein kinases required for cell cycle progression, *Nature*, 432(7020), 980–987.

**Billingsley, M.L., Yun, J., Reese, B.E., Davidson, C.E., Buck-Koehntop, B.A. & Veglia, G.**, 2006, Functional and structural properties of stannin: roles in cellular growth, selective toxicity, and mitochondrial responses to injury, *Journal of Cellular Biochemistry*, 98(2), 243–250.

**Björklund, M., Taipale, M., Varjosalo, M., Saharinen, J., Lahdenperä, J. & Taipale, J**., 2006, Identification of pathways regulating cell size and cell-cycle progression by RNAi, *Nature*, 439(7079), 1009–1013.

**Blaszczyk, J., Tropea, J.E., Bubunenko, M., Routzahn, K.M., Waugh, D.S., Court, D.L. & Ji, X.**, 2001, Crystallographic and modeling studies of RNase III suggest a mechanism for double-stranded RNA cleavage, *Structure*, 9(12), 1225–1236.

**Botelho, R.J. & Grinstein, S.**, 2011, Phagocytosis, *Current Biology*, 21(14), R533–538.

**Boutros, M. & Ahringer, J.**, 2008, The art and design of genetic screens: RNA interference, *Nature Reviews Genetics*, 9(7), 554–566.

**Boutros, M., Kiger, A.A., Armknecht, S., Kerr, K., Hild, M., Koch, B., Haas, S.A., Paro, R., Perrimon, N. & Heidelberg Fly Array Consortium,** 2004, Genome-wide RNAi analysis of growth and viability in Drosophila cells, *Science*, 303(5659), 832–835.

**Brar, G.A. & Weissman, J.S.**, 2015, Ribosome profiling reveals the what, when, where and how of protein synthesis, *Nature Reviews Molecular Cell Biology*, 16(11), 651–664.

**Brunner, E., Ahrens, C.H., Mohanty, S., Baetschmann, H., Loevenich, S., Potthast, F., Deutsch, E.W., Panse, C., de Lichtenberg, U., Rinner, O., Lee, H., Pedrioli, P.G., Malmstrom, J., Koehler, K., Schrimpf, S., Krijgsveld, J., Kregenow, F., Heck, A.J., Hafen, E., Schlapbach, R. & Aebersold, R.,** 2007, A high-quality catalog of the Drosophila melanogaster proteome, *Nature Biotechnology*, 25(5), 576–583.

**Burman, J.L., Itsara, L.S., Kayser, E.B., Suthammarak, W., Wang, A.M., Kaeberlein, M., Sedensky, M.M., Morgan, P.G. & Pallanck, L.J.,** 2014, A Drosophila model of mitochondrial disease caused by a complex I mutation that uncouples proton pumping from electron transfer, *Disease Models & Mechanisms*, 7(10), 1165–1174.

**de Castro, I.P., Martins, L.M. & Loh, S.H.,** 2011, Mitochondrial quality control and Parkinson's disease: a pathway unfolds, *Molecular Neurobiology*, 43(2), 80–86.

**Cermelli, S., Guo, Y., Gross, S.P. & Welte, M.A.,** 2006, The lipid-droplet proteome reveals that droplets are a protein-storage depot, *Current Biology*, 16(18), 1783–1795.

**Cevallos, R.C. & Sarnow, P.,** 2010, Temperature protects insect cells from infection by cricket paralysis virus, *Journal of Virology*, 84(3), 1652–1655.

**Charroux, B. & Royet, J.,** 2009, Elimination of plasmatocytes by targeted apoptosis reveals their role in multiple aspects of the Drosophila immune response, *Proceedings of the National Academy of Sciences of the United States of America*, 106(24), pp. 9797-802.

**Chen, H. & Chan, D.C.,** 2009, Mitochondrial dynamics--fusion, fission, movement, and mitophagy--in neurodegenerative diseases, *Human Molecular Genetics*, 18(R2), R169–R176.

**Cherbas, L., Moss, R. & Cherbas, P.,** 1994, Transformation techniques for Drosophila cell lines, *Methods in Cell Biology*, 44, 161–179.

**Cherbas, L., Willingham, A., Zhang, D., Yang, L., Zou, Y., Eads, B.D., Carlson, J.W., Landolin, J.M., Kapranov, P., Dumais, J., Samsonova, A., Choi, J.H., Roberts, J., Davis, C.A., Tang, H., van Baren, M.J., Ghosh, S., Dobin, A., Bell, K., Lin, W., Langton, L., Duff, M.O., Tenney, A.E., Zaleski, C., Brent, M.R., Hoskins, R.A., Kaufman, T.C., Andrews, J., Graveley, B.R., Perrimon, N., Celniker, S.E., Gingeras, T.R. & Cherbas, P.,** 2011, The transcriptional diversity of 25 Drosophila cell lines, *Genome research*, 21(2), 301-14.

**Chew, G.L., Pauli, A., Rinn, J.L., Regev, A., Schier, A.F. & Valen, E.,** 2013, Ribosome profiling reveals resemblance between long non-coding RNAs and 5' leaders of coding RNAs, *Development*, 140(13), 2828–2834.

**Child, S.J., Miller, M.K. & Geballe, A.P.,** 1999, Translational control by an upstream open reading frame in the HER-2/neu transcript, *The Journal of Biological Chemistry*, 274(34), 24335–24341.

**Cho, N.K., Keyes, L., Johnson, E., Heller, J., Ryner, L., Karim, F. & Krasnow, M.A.,** 2002, Developmental control of blood cell migration by the Drosophila VEGF pathway, *Cell*, 108(6), 865–876.

**Chung, Y.-T. & Keller, E.B.,** 1990, Positive and negative regulatory elements mediating transcription from the Drosophila melanogaster actin 5C distal promoter, *Molecular and Cellular Biology*, 10(12), 6172–6180.

**Clark, A.G., Glanowski, S., Nielsen, R., Thomas, P.D., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X., White, T.J., Sninsky, J.J., Adams, M.D. & Cargill, M.,** 2003, Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios, *Science*, 302(5652), 1960–1963.

**Clemens, J.C., Worby, C.A., Simonson-Leff, N., Muda, M., Maehama, T., Hemmings, B.A. & Dixon, J.E.,** 2000, Use of double-stranded RNA interference in Drosophila cell lines to dissect signal transduction pathways, *Proceedings of the National Academy of Sciences of the United States of America*, 97(12), 6499-64503.

**Colgan, D.F. & Manley, J.L.,** 1997, Mechanism and regulation of mRNA polyadenylation, *Genes & Development*, 11(21), 2755–2766.

**Consortium, I.H.G.S.,** 2004, Finishing the euchromatic sequence of the human genome, *Nature*, 431(7011), 931-45.

**Costa, F.F.,** 2007, Non-coding RNAs: lost in translation? *Gene*, 386(1–2), 1–10.

**Couso, J.P. & Bishop, S.A.,** 1998, Proximo-distal development in the legs of Drosophila, *International Journal of Developmental Biology*, 42(3 Spec No), 345–352.

**Cox, R.T. & Spradling, A.C.,** 2009, Clueless, a conserved Drosophila gene required for mitochondrial subcellular localization, interacts genetically with parkin, *Disease Models & Mechanisms*, 2(9–10), pp. 490–499.

**Crozatier, M. & Meister, M.,** 2007, Drosophila haematopoiesis, *Cellular Microbiology*, 9(5), 1117–1126.

**Das, S., Yu, L., Gaitatzes, C., Rogers, R., Freeman, J., Bienkowska, J., Adams, R.M., Smith, T.F. & Lindelien, J.,** 1997, Biology's new Rosetta stone, *Nature*, 385(6611), 29–30.

**Davidson, C.E., Reese, B.E., Billingsley, M.L. & Yun, J.K.,** 2004, Stannin, a protein that localizes to the mitochondria and sensitizes NIH-3T3 cells to trimethyltin and dimethyltin toxicity, *Molecular Pharmacology*, 66(4), 855–863.

**Davidson, C.E., Reese, B.E., Billingsley, M.L. & Yun, J.K.,** 2005a, The protein stannin binds 14-3-3eta and modulates mitogen-activated protein kinase signaling, *Molecular Brain Research*, 138(2), 256–263.

**Davidson, C.E., Reese, B.E., Billingsley, M.L. & Yun, J.K.,** 2005b, The protein stannin binds 14-3-3eta and modulates mitogen-activated protein kinase signaling, *Molecular Brain Research*, 138(2), 256–263.

**Day, D.A. & Tuite, M.F.,** 1998, Post-transcriptional gene regulatory mechanisms in eukaryotes: an overview, *The Journal of Endocrinology*, 157(3), 361–371.

**Defaye, A., Evans, I., Crozatier, M., Wood, W., Lemaitre, B. & Leulier, F.,** 2009, Genetic ablation of Drosophila phagocytes reveals their contribution to both development and resistance to bacterial infection, *Journal of Innate Immunity*, 1(4), 322–334.

**Dejneka, N.S., Patanow, C.M., Polavarapu, R., Toggas, S.M., Krady, J.K. & Billingsley, M.L.,** 1997, Localization and characterization of stannin: relationship to cellular sensitivity to organotin compounds, *Neurochemistry International*, 31(6), 801–815.

**Demir, E. & Dickson, B.J.,** 2005, fruitless splicing specifies male courtship behavior in Drosophila, *Cell*, 121(5), 785–794.

**Desagher, S. & Martinou, J.-C.,** 2000, Mitochondria as the central control point of apoptosis, *Trends in Cell Biology*, 10(9), 369–377.

**Desiere,F., Deutsch, E.W., King, N.L., Nesvizhskii, A.I., Mallick, P., Eng, J., Chen, S., Eddes, J., Loevenich, S.N., Aebersold, R.,** 2007, *The PeptideAtlas Project* Nucleic Acids Research. 34, D655-D658

**Diawara, M.R., Hue, C., Wilder, S.P., Venteclef, N., Aron-Wisnewsky, J., Scott, J., Clément, K., Gauguier, D. & Calderari, S.,** 2014, Adaptive expression of microRNA-125a in adipose tissue in response to obesity in mice and men, *PloS one*, 9(3), e91375.

**Dinger, M.E., Pang, K.C., Mercer, T.R. & Mattick, J.S.,** 2008, Differentiating protein-coding and noncoding RNA: challenges and ambiguities, *PLoS Computational Biology*, 4(11), e1000176.

**Duncan, C.D. & Mata, J.,** 2014, The translational landscape of fission-yeast meiosis and sporulation, *Nature Structural & Molecular Biology*, 21(7), 641–647.

**Dunn, J.G., Foo, C.K., Belletier, N.G., Gavis, E.R. & Weissman, J.S.,** 2013, Ribosome profiling reveals pervasive and regulated stop codon readthrough in Drosophila melanogaster, *eLife*, 2, e01179.

**Dunn, K.W., Kamocka, M.M. & McDonald, J.H.,** 2011, A practical guide to evaluating colocalization in biological microscopy, *American journal of physiology. Cell Physiology*, 300(4), C723–C742.

**Ebina, M., Tsuruta, F., Katoh, M.C., Kigoshi, Y., Someya, A. & Chiba, T.,** 2013, Myeloma overexpressed 2 (Myeov2) regulates L11 subnuclear localization through Nedd8 modification, *PloS one*, 8(6), e65285.

**Edgar, B.A., Sprenger, F., Duronio, R.J., Leopold, P. & O'Farrell, P.H.,** 1994, Distinct molecular mechanism regulate cell cycle timing at successive stages of Drosophila embryogenesis, *Genes & Development*, 8(4), 440–452.

**Eggert, U.S., Kiger, A.A., Richter, C., Perlman, Z.E., Perrimon, N., Mitchison, T.J. & Field, C.M.,** 2004, Parallel chemical genetic and genome-wide RNAi screens identify cytokinesis inhibitors and targets, *PLoS Biology*, 2(12), e379.

**Einhauer, A. & Jungbauer, A.,** 2001, The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins, *Journal of Biochemical and Biophysical Methods*, 49(1–3), 455–465.

**Elwell, C. & Engel, J.N.,** 2005, Drosophila melanogaster S2 cells: a model system to study Chlamydia interaction with host cells, *Cellular Microbiology*, 7(5), 725–739.

**Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H.,** 2007, Locating proteins in the cell using TargetP, SignalP and related tools, *Nature Protocols*, 2(4), 953–971.

**Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G.,** 2000, Predicting subcellular localization of proteins based on their N-terminal amino acid sequence, *Journal of Molecular Biology*, 300(4), 1005–1016.

**Ernster, L. & Schatz, G.,** 1981, Mitochondria: a historical review, *The Journal of Cell Biology*, 91(3), 227s-255s.

**Evans, E.K., Kuwana, T., Strum, S.L., Smith, J.J., Newmeyer, D.D. & Kornbluth, S.,** 1997, Reaper-induced apoptosis in a vertebrate system, *The EMBO journal*, 16(24), pp. 7372-81.

**Evans, I.R., Hu, N., Skaer, H. & Wood, W.,** 2010a, Interdependence of macrophage migration and ventral nerve cord development in Drosophila embryos, *Development*, 137(10), 1625–1633.

**Evans, I.R., Zanet, J., Wood, W. & Stramer, B.M.,** 2010b, Live imaging of Drosophila melanogaster embryonic hemocyte migrations, *Journal of Visualized Experiments: JoVE,* (36).

**Fassler, J., Landsman, D., Acharya, A., Moll, J.R., Bonovich, M. & Vinson, C.,** 2002, B-ZIP proteins encoded by the Drosophila genome: evaluation of potential dimerization partners, *Genome Research*, 12(8), 1190–1200.

**Federico, A., Cardaioli, E., Da Pozzo, P., Formichi, P., Gallus, G.N. & Radi, E.,** 2012, Mitochondria, oxidative stress and neurodegeneration, *Journal of Neurological Sciences*, 322(1–2), pp. 254–262.

**Fickett, J.W.,** 1995, ORFs and genes: how strong a connection? *Journal of Computational Biology*, 2(1),117–123.

**Finkel, T.,** 2015, The metabolic regulation of aging, *Nat Med*, 21(12), pp. 1416-23.

**Fire, A.,** 1999, RNA-triggered gene silencing, *Trends in Genetics*, 15(9), 358–363.

**Fossett, N., Tevosian, S.G., Gajewski, K., Zhang, Q., Orkin, S.H. & Schulz, R.A.,** 2001, The Friend of GATA proteins U-shaped, FOG-1, and FOG-2 function as

negative regulators of blood, heart, and eye development in Drosophila, *Proceedings of the National Academy of Sciences*, 98(13), 7342–7347.

**Franc, N.C.,** 1999, Requirement for Croquemort in Phagocytosis of Apoptotic Cells in Drosophila, *Science*, 284(5422), 1991–1994.

**Franc, N.C., Dimarcq, J.-L., Lagueux, M., Hoffmann, J. & Ezekowitz, R.A.B.,** 1996, Croquemort, a novel Drosophila hemocyte/macrophage receptor that recognizes apoptotic cells, *Immunity*, 4(5), 431–443.

**Freel, C.D., Richardson, D.A., Thomenius, M.J., Gan, E.C., Horn, S.R., Olson, M.R. & Kornbluth, S.,** 2008, Mitochondrial localization of Reaper to promote inhibitors of apoptosis protein degradation conferred by GH3 domain-lipid interactions, *The Journal of Biological Chemistry*, 283(1), 367–379.

**Frieden, M., James, D., Castelbou, C., Danckaert, A., Martinou, J.C. & Demaurex, N.,** 2004, Ca(2+) homeostasis during mitochondrial fragmentation and perinuclear clustering induced by hFis1, *The Journal of Biological Chemistry*, 279(21), 22704–2214.

**Frith, M.C., Forrest, A.R., Nourbakhsh, E., Pang, K.C., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., Bailey, T.L. & Grimmond, S.M.,** 2006, The abundance of short proteins in the mammalian proteome, *Plos Genetics*, 2(4), 515–528.

**Fu, H., Subramanian, R.R. & Masters, S.C.,** 2000, 14-3-3 proteins: structure, function, and regulation, *Annual Review of Pharmacology and Toxicology*, 40(1), 617–647.

**Fullard, J.F., Kale, A. & Baker, N.E.,** 2009, Clearance of apoptotic corpses, *Apoptosis*, 14(8), 1029–1037.

**Galindo, M.I. & Couso, J.P.,** 2000, Intercalation of cell fates during tarsal development in Drosophila, *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 22(9), 777–780.

**Galindo, M.I., Bishop, S.A. & Couso, J.P.,** 2005, Dynamic EGFR-Ras signalling in Drosophila leg development, *Developmental Dynamics*, 233(4), 1496–1508.

**Galindo, M.I., Bishop, S.A., Greig, S. & Couso, J.P.,** 2002, Leg patterning driven by proximal-distal interactions and EGFR signaling, *Science*, 297(5579), 256–259.

**Galindo, M.I., Pueyo, J.I., Fouix, S., Bishop, S.A. & Couso,** J.P., 2007, Peptides encoded by short ORFs control development and define a new eukaryotic gene family, *PLoS Biology*, 5(5), e106.

**Gandre-Babbe, S. & van der Bliek, A.M.,** 2008, The novel tail-anchored membrane protein Mff controls mitochondrial and peroxisomal fission in mammalian cells, *Molecular Biology of the Cell*, 19(6), 2402–2412.

**Garen, A., Kauvar, L. & Lepesant, J.-A.,** 1977, Roles of ecdysone in Drosophila development, *Proceedings of the National Academy of Sciences*, 74(11), 5099–5103.

**Garesse, R.,** 1988, Drosophila melanogaster mitochondrial DNA: gene organization and evolutionary considerations, *Genetics*, 118(4), 649–663.

**Gershman, B., Puig, O., Hang, L., Peitzsch, R.M., Tatar, M. & Garofalo, R.S.,** 2007, High-resolution dynamics of the transcriptional response to nutrition in Drosophila: a key role for dFOXO, *Physiological Genomics*, 29(1), 24–34.

**Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C. & Johnston, M.,** 1996, Life with 6000 genes, *Science*, 274(5287), 546–567.

**Gomes, L.C., Di Benedetto, G. & Scorrano, L.,** 2011, During autophagy mitochondria elongate, are spared from degradation and sustain cell viability, *Nature Cell Biology*, 13(5), 589–598.

**Goto, A., Kadowaki, T. & Kitagawa, Y., 2003,** Drosophila hemolectin gene is expressed in embryonic and larval hemocytes and its knock down causes bleeding defects, *Developmental Biology*, 264(2), 582–591.

**Goyal, G., Fell, B., Sarin, A., Youle, R.J. & Sriram, V.,** 2007, Role of mitochondrial remodeling in programmed cell death in Drosophila melanogaster, *Developmental Cell*, 12(5), 807–816.

**Greene, J.C., Whitworth, A.J., Kuo, I., Andrews, L.A., Feany, M.B. & Pallanck, L.J.,** 2003, Mitochondrial pathology and apoptotic muscle degeneration in Drosophila parkin mutants, *Proceedings of the National Academy of Sciences*, 100(7), 4078–4083.

**Grünewald, A., Voges, L., Rakovic, A., Kasten, M., Vandebona, H., Hemmelmann, C., Lohmann, K., Orolicki, S., Ramirez, A., Schapira, A.H., Pramstaller, P.P., Sue, C.M. & Klein, C.,** 2010, Mutant Parkin impairs mitochondrial function and morphology in human fibroblasts, *PloS one*, 5(9), e12962.

**Haga, N., Fujita, N. & Tsuruo, T.,** 2003, Mitochondrial aggregation precedes cytochrome c release from mitochondria during apoptosis, *Oncogene*, 22(36), 5579–5585.

**Haga, N., Fujita, N. & Tsuruo, T.,** 2005, Involvement of mitochondrial aggregation in arsenic trioxide (As2O3)-induced apoptosis in human glioblastoma cells, *Cancer Science*, 96(11), 825–833.

**Haley, B., Hendrix, D., Trang, V. & Levine, M.,** 2008, A simplified miRNA-based gene silencing method for Drosophila melanogaster, *Developmental Biology*, 321(2), 482–490.

**Hamalainen-Laanaya, H.K. & Orloff, M.S.,** 2012, Analysis of cell viability using time-dependent increase in fluorescence intensity, *Analytical Biochemistry*, 429(1), 32–38.

**Hammond, S.M., Caudy, A.A. & Hannon, G.J.,** 2001, Post-transcriptional gene silencing by double-stranded RNA, *Nature Reviews Genetics*, 2(2), 110–119.

**Hanada, K., Zhang, X., Borevitz, J.O., Li, W.H. & Shiu, S.H.,** 2007, A large number of novel coding small open reading frames in the intergenic regions of the Arabidopsis thaliana genome are transcribed and/or under purifying selection, *Genome Research*, 17(5), 632–640.

**Hartenstein, V.,** 1985, *The Embryonic Development of Drosophila melanogaster,* Springer-Verlag, Berlin; New York.

**Hatchell, E.C., Colley, S.M., Beveridge, D.J., Epis, M.R., Stuart, L.M., Giles, K.M., Redfern, A.D., Miles, L.E., Barker, A., MacDonald, L.M., Arthur, P.G., Lui, J.C., Golding, J.L., McCulloch, R.K., Metcalf, C.B., Wilce, J.A., Wilce, M.C., Lanz, R.B., O'Malley, B.W. & Leedman, P.J.,** 2006, SLIRP, a small SRA binding protein, is a nuclear receptor corepressor, *Molecular Cell*, 22(5), 657–668.

**Hoffmann, J.A.,** 1999, Phylogenetic Perspectives in Innate Immunity, *Science*, 284(5418), 1313–1318.

**Holz, A., Bossinger, B., Strasser, T., Janning, W. & Klapper, R.,** 2003, The two origins of hemocytes in Drosophila, *Development*, 130(20), 4955–4962.

**Horrevoets, A.J., Fontijn, R.D., van Zonneveld, A.J., de Vries, C.J., ten Cate, J.W. & Pannekoek, H.,** 1999, Vascular Endothelial Genes That Are Responsive to Tumor Necrosis Factor In Vitro Are Expressed in Atherosclerotic Lesions, Including Inhibitor of Apoptosis Protein-1, Stannin, and Two Novel Genes, *Blood*, 93(10), 3418–3431.

**Humphries, W.H. & Payne, C.K.,** 2012, Imaging lysosomal enzyme activity in live cells using self-quenched substrates, *Analytical Biochemistry*, 424(2), 178–183.

**Huotari, J. & Helenius, A.,** 2011, Endosome maturation, *The EMBO Journal*, 30(17), 3481–3500.

**Inagaki, S., Numata, K., Kondo, T., Tomita, M., Yasuda, K., Kanai, A., Kageyama, Y.,** 2005, Identification and expression analysis of putative mRNA-like non-coding RNA in Drosophila, *Genes to Cells,*10(12),1163-73.

**Ingolia, N.T., Brar, G.A., Stern-Ginossar, N., Harris, M.S., Talhouarne, G.J., Jackson, S.E., Wills, M.R. & Weissman, J.S.,** 2014, Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes, *Cell Reports*, 8(5), 1365–1379.

**Ingolia, N.T., Ghaemmaghami, S., Newman, J.R. & Weissman, J.S.,** 2009, Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling, *Science*, 324(5924), 218–223.

**Ingolia, N.T., Lareau, L.F. & Weissman, J.S.,** 2011, Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes, *Cell*, 147(4), 789–7802.

**Ishihara, N., Jofuku, A., Eura, Y. & Mihara, K.,** 2003, Regulation of mitochondrial morphology by membrane potential, and DRP1-dependent division and FZO1-dependent fusion reaction in mammalian cells, *Biochemical and Biophysical Research Communications*, 301(4), 891–898.

**Itoh, K., Nakamura, K., Iijima, M. & Sesaki, H.,** 2013, Mitochondrial dynamics in neurodegeneration, *Trends in Cell Biology*, 23(2), 64–71.

**Ivatt, R.M., Sanchez-Martinez, A., Godena, V.K., Brown, S., Ziviani, E. & Whitworth, A.J.,** 2014, Genome-wide RNAi screen identifies the Parkinson disease GWAS risk locus SREBF1 as a regulator of mitophagy, *Proceedings of the National Academy of Sciences of the United States of America*, 111(23), 8494–8499.

**James, A.M., Collins, Y., Logan, A. & Murphy, M.P.,** 2012, Mitochondrial oxidative stress and the metabolic syndrome, *Trends in Endocrinology and Metabolism: TEM*, 23(9), 429–434.

**Jean, S., Cox, S., Schmidt, E.J., Robinson, F.L. & Kiger, A.,** 2012, Sbf/MTMR13 coordinates PI (3) P and Rab21 regulation in endocytic control of cellular remodeling, *Molecular Biology of the Cell*, 23(14), 2723–2740.

**Ji, B., Kim, M., Higa, K.K. & Zhou, X.,** 2015, Boymaw, Overexpressed in Brains With Major Psychiatric Disorders, May Encode a Small Protein to Inhibit

Mitochondrial Function and Protein Translation, *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 168(4), 284–295.

**Jiang, C., Baehrecke, E.H. & Thummel, C.S.,** 1997, Steroid regulated programmed cell death during Drosophila metamorphosis, *Development, England)*, 124(22), pp. 4673–4683.

**Jornayvaz, F.R. & Shulman, G.I.,** 2010, Regulation of mitochondrial biogenesis, *Essays in Biochemistry*, 47, 69–84.

**Jung, S.-H., Evans, C.J., Uemura, C. & Banerjee, U.,** 2005, The Drosophila lymph gland as a developmental model of hematopoiesis, *Development*, 132(11), pp. 2521–2533.

**Karbowski, M. & Youle, R.J.,** 2003, Dynamics of mitochondrial morphology in healthy cells and during apoptosis, *Cell Death & Differentiation*, 10(8), 870–880.

**Kastenmayer, J.P., Ni, L., Chu, A., Kitchen, L.E., Au, W.-C., Yang, H., Carter, C.D., Wheeler, D., Davis, R.W., Boeke, J.D., Snyder, M.A. & Basrai, M.A.,** 2006, Functional genomics of genes with small open reading frames (sORFs) in S. cerevisiae, *Genome Research*, 16(3), 365–373.

**Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. & Sternberg, M.J.,** 2015, The Phyre2 web portal for protein modeling, prediction and analysis, *Nature Protocols*, 10(6), 845–858.

**Kiger, A.A., Baum, B., Jones, S., Jones, M.R., Coulson, A., Echeverri, C. & Perrimon, N.,** 2003, A functional genomic analysis of cell morphology using RNA interference, *Journal of Biology*, 2(4), 27.

**Kondo, T., Hashimoto, Y., Kato, K., Inagaki, S., Hayashi, S. & Kageyama, Y.,** 2007, Small peptide regulators of actin-based cell morphogenesis encoded by a dicistronic mRNA, *Nature Cell Biology*, 9(6), 660–665.

**Kondylis, V., Tang, Y., Fuchs, F., Boutros, M. & Rabouille, C.,** 2011, Identification of ER proteins involved in the functional organisation of the early secretory pathway in Drosophila cells by a targeted RNAi screen, *PloS one*, 6(2), e17173.

**Kozlova, T. & Thummel, C.S.,** 2003, Essential roles for ecdysone signaling during Drosophila mid-embryonic development, *Science*, 301(5641), 1911–1914.

**Krauss, S.,** 2001, Mitochondria: Structure and role in respiration, *eLS*.

**Ladoukakis, E., Pereira, V., Magny, E.G., Eyre-Walker, A. & Couso, J.P.,** 2011, Hundreds of putatively functional small open reading frames in Drosophila, *Genome Biology*, 12(11), R118.

**Lane, N. & Martin, W.,** 2010, The energetics of genome complexity, *Nature*, 467(7318), 929–934.

**Lanot, R., Zachary, D., Holder, F. & Meister, M.,** 2001, Postembryonic hematopoiesis in Drosophila, *Developmental Biology*, 230(2), 243–257.

**Le, T., Liang, Z., Patel, H., Yu, M.H., Sivasubramaniam, G., Slovitt, M., Tanentzapf, G., Mohanty, N., Paul, S.M., Wu, V.M. & Beitel, G.J.,** 2006, A new family of Drosophila balancer chromosomes with a w- dfd-GMR yellow fluorescent protein marker, *Genetics*, 174(4), 2255–2257.

**Lebestky, T.,** 2000, Specification of Drosophila Hematopoietic Lineage by Conserved Transcription Factors, *Science*, 288(5463), 146–149.

**Lee, C., Yen, K. & Cohen, P.,** 2013, Humanin: a harbinger of mitochondrial-derived peptides? *Trends in Endocrinology and Metabolism: TEM*, 24(5), 222–228.

**Lee, C., Zeng, J., Drew, B.G., Sallam, T., Martin-Montalvo, A., Wan, J., Kim, S.J., Mehta, H., Hevener, A.L., de Cabo, R. & Cohen, P.,** 2015, The mitochondrial-derived peptide MOTS-c promotes metabolic homeostasis and reduces obesity and insulin resistance, *Cell Metabolism*, 21(3), 443–454.

**Lee, D.F., Chen, C.C., Hsu, T.A. & Juang, J.L.,** 2000, A baculovirus superinfection system: efficient vehicle for gene transfer into Drosophila S2 cells, *Journal of Virology*, 74(24), 11873–11880.

**Lee, J.H., Cho, K.S., Lee, J., Yoo, J., Lee, J. & Chung, J.,** 2001, Diptericin-like protein: an immune response gene regulated by the anti-bacterial gene induction pathway in Drosophila, *Gene*, 271(2), 233–238.

**Lee, Y.K., Hur, W., Lee, S.W., Hong, S.W., Kim, S.W., Choi, J.E. & Yoon, S.K.,** 2014, Knockdown of 14-3-3ζ enhances radiosensitivity and radio-induced apoptosis in CD133(+) liver cancer stem cells, *Experimental & Molecular Medicine* 46, e77.

**Lemaitre, B. & Hoffmann, J.,** 2007, The host defense of Drosophila melanogaster, *Annual Review of Immunology*, 25, 697–743.

**Lemaitre, B., Reichhart, J.M. & Hoffmann, J.A.,** 1997, Drosophila host defense: differential induction of antimicrobial peptide genes after infection by various classes of microorganisms, *Proceedings of the National Academy of Sciences of the United States of America*, 94(26), 14614–14619.

**Li, J.J., Huang, H., Bickel, P.J. & Brenner, S.E.,** 2014, Comparison of D. melanogaster and C. elegans developmental stages, tissues, and cells by modENCODE RNA-seq data, *Genome Research*, 24(7), 1086–1101.

**Li, Y. & Kiledjian, M.,** 2010, Regulation of mRNA decapping, *Wiley Interdisciplinary Reviews: RNA*, 1(2), 253–265.

**Lin, M.T. & Beal, M.F.,** 2006, Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases, *Nature*, 443(7113), 787–795.

**Lindbo, J.A., Silva-Rosales, L., Proebsting, W.M. & Dougherty, W.G.,** 1993, Induction of a Highly Specific Antiviral State in Transgenic Plants: Implications for Regulation of Gene Expression and Virus Resistance, *The Plant Cell*, 5(12), 1749–1759.

**Lindmo, K. & Stenmark, H.,** 2006, Regulation of membrane traffic by phosphoinositide 3-kinases, *Journal of Cell Science*, 119(Pt 4), 605–614.

Lindquist, S., 1984, Heat shock--a comparison of Drosophila and yeast, *Journal of Embryology and Experimental Morphology*, 83 Suppl, 147–161.

**Lipardi, C., Wei, Q. & Paterson, B.M.,** 2001, RNAi as random degradative PCR: siRNA primers convert mRNA into dsRNAs that are degraded to generate new siRNAs, *Cell*, 107(3), 297–307.

**Liu, Q., Rand, T.A., Kalidas, S., Du, F., Kim, H.-E., Smith, D.P. & Wang, X.,** 2003, R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway, *Science*, 301(5641), 1921–1925.

**Liu, X., Kim, C.N., Yang, J., Jemmerson, R. & Wang, X.,** 1996, Induction of apoptotic program in cell-free extracts: requirement for dATP and cytochrome c, *Cell*, 86(1), 147–157.

**Lodish, H.,** 2008, *Molecular Cell Biology,* Macmillan,.

**Luzio, J.P., Rous, B.A., Bright, N.A., Pryor, P.R., Mullock, B.M. & Piper, R.C.,** 2000, Lysosome-endosome fusion and lysosome biogenesis, *Journal of Cell Science*, 113 (Pt 9), 1515–1524.

**Ma, F., Xu, S., Liu, X., Zhang, Q., Xu, X., Liu, M., Hua, M., Li, N., Yao, H. & Cao, X.,** 2011, The microRNA miR-29 controls innate and adaptive immune responses to intracellular bacterial infection by targeting interferon-γ, *Nature Immunology*, 12(9), 861–89.

**Mackowiak, S.D., Zauber, H., Bielow, C., Thiel, D., Kutz, K., Calviello, L., Mastrobuoni, G., Rajewsky, N., Kempa, S., Selbach, M. & Obermayer, B.,** 2015, Extensive identification and analysis of conserved small ORFs in animals, *Genome Biology*, 16, 179.

**Magny, E.G., Pueyo, J.I., Pearl, F.M., Cespedes, M.A., Niven, J.E., Bishop, S.A. & Couso, J.P.,** 2013, Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames, *Science*, 341(6150), 1116–1120.

**McBride, H.M., Neuspiel, M. & Wasiak, S.,** 2006, Mitochondria: more than just a powerhouse, *Current Biology*, 16(14), R551–R560.

**de Meis, L., Ketzer, L.A., da Costa, R.M., de Andrade, I.R. & Benchimol, M.,** 2010, Fusion of the endoplasmic reticulum and mitochondrial outer membrane in rats brown adipose tissue: activation of thermogenesis by Ca2+, *PloS one*, 5(3), e9439.

**Moreira, C.G., Regan, J.C., Zaidman-Remy, A., Jacinto, A. & Prag, S.,** 2011, Drosophila hemocyte migration: an in vivo assay for directional cell migration, *Methods in Molecular Biology*, 769, 249–260.

**Misra, S., Crosby, M.A., Mungall, C.J., Matthews, B.B., Campbell, K.S., Hradecky, P., Huang, Y., Kaminker, J.S., Millburn, G.H., Prochnik, S.E., Smith, C.D., Tupy, J.L., Whitfied, E.J., Bayraktaroglu, L., Berman, B.P., Bettencourt, B.R., Celniker, S.E., de Grey, A.D., Drysdale, R.A., Harris, N.L., Richter, J., Russo, S., Schroeder, A.J., Shu, S.Q., Stapleton, M., Yamada, C., Ashburner, M., Gelbart, W.M., Rubin, G.M. & Lewis, S.E.,** 2002, Annotation of the Drosophila melanogaster euchromatic genome: a systematic review, *Genome biology*, 3(12), p. RESEARCH0083.

**Mumtaz, M.A. & Couso, J.P.,** 2015, Ribosomal profiling adds new coding sequences to the proteome, *Biochemical Society Transactions*, 43(6), 1271–1276.

**Neal, C.L., Yao, J., Yang, W., Zhou, X., Nguyen, N.T., Lu, J., Danes, C.G., Guo, H., Lan, K.H., Ensor, J., Hittelman, W., Hung, M.C. & Yu, D.,** 2009, 14-3-3eta overexpression defines high risk for breast cancer recurrence and promotes cancer cell survival, *Cancer Research*, 69(8), 3425–3432.

**Nelson, R.E., Fessler, L.I., Takagi, Y., Blumberg, B., Keene, D.R., Olson, P.F., Parker, C.G. & Fessler, J.H.,** 1994, Peroxidasin: a novel enzyme-matrix protein of Drosophila development, *The EMBO Journal*, 13(15), 3438.

**Neupert, W. & Herrmann, J.M.,** 2007, Translocation of proteins into mitochondria, *Annual Review of Biochemistry*, 76, 723–749.

**Nguyen, H.T. & Frasch, M.,** 2006, MicroRNAs in muscle differentiation: lessons from Drosophila and beyond, *Current Opinion in Genetics and Development*, 16(5), 533–539.

**Norgate, M., Southon, A., Zou, S., Zhan, M., Sun, Y., Batterham, P. & Camakaris, J.,** 2007, Copper homeostasis gene discovery in Drosophila melanogaster, *Biometals*, 20(3–4), 683–697.

**Norris, K.L., Hao, R., Chen, L.F., Lai, C.H., Kapur, M., Shaughnessy, P.J., Chou, D., Yan, J., Taylor, J.P., Engelender, S., West, A.E., Lim, K.L. & Yao, T.P.,** 2015, Convergence of Parkin, PINK1, and α-Synuclein on Stress-induced Mitochondrial Morphological Remodeling, *The Journal of Biological Chemistry*, 290(22), 13862–13874.

**Nunnari, J. & Suomalainen, A.,** 2012, Mitochondria: in sickness and in health, *Cell*, 148(6), 1145–1159.

**Nwaneshiudu, A., Kuschal, C., Sakamoto, F.H., Anderson, R.R., Schwarzenberger, K. & Young, R.C.,** 2012, Introduction to confocal microscopy, *Journal of Investigative Detmatology*, 132(12), e3.

**Nykänen, A., Haley, B. & Zamore, P.D.,** 2001, ATP requirements and small interfering RNA structure in the RNA interference pathway, *Cell*, 107(3), 309–321.

**Okamoto, M., Higuchi-Takeuchi, M., Shimizu, M., Shinozaki, K. & Hanada, K.,** 2014, Substantial expression of novel small open reading frames in Oryza sativa, *Plant Signaling & Behavior*, 9(2), e27848.

**Okatsu, K., Saisho, K., Shimanuki, M., Nakada, K., Shitara, H., Sou, Y.S., Kimura, M., Sato, S., Hattori, N., Komatsu, M., Tanaka, K. & Matsuda, N.,** 2010, p62/SQSTM1 cooperates with Parkin for perinuclear clustering of depolarized mitochondria, *Genes to Cells*, 15(8), 887–900.

**Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R. & Suzuki, H.,** 2002, Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs, *Nature*, 420(6915), 563–573.

**Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., and consortium,** 2004, Complete sequencing and characterization of 21,243 full-length human cDNAs, *Nature Genetics*, 36(1), 40–45.

**Park, J., Kim, Y. & Chung, J.,** 2009, Mitochondrial dysfunction and Parkinson's disease genes: insights from Drosophila, *Disease Models & Mechanisms*, 2(7–8), 336–340.

**Park, J., Kim, Y., Choi, S., Koh, H., Lee, S.H., Kim, J.M. & Chung, J.,** 2010, Drosophila Porin/VDAC affects mitochondrial morphology, *PloS one*, 5(10), e13151.

**Pollard, K.S., Salama, S.R., Lambert, N., Lambot, M.A., Coppens, S., Pedersen, J.S., Katzman, S., King, B., Onodera, C., Siepel, A., Kern, A.D., Dehay, C., Igel, H., Ares, M., Vanderhaeghen, P. & Haussler, D.,** 2006, An RNA gene expressed during cortical development evolved rapidly in humans, *Nature*, 443(7108), 167–172.

**Pucci, B., Bertani, F., Karpinich, N.O., Indelicato, M., Russo, M.A., Farber, J.L. & Tafani, M.,** 2008, Detailing the role of Bax translocation, cytochrome c release, and perinuclear clustering of the mitochondria in the killing of HeLa cells by TNF, *Journal of Cellular Physiology*, 217(2), 442–449.

**Pueyo, J.I. & Couso, J.P.,** 2005, Parallels between the proximal-distal development of vertebrate and arthropod appendages: homology without an ancestor? *Current Opinion in Genetics & Development*, 15(4), 439–446.

**Pueyo, J.I. & Couso, J.P.,** 2008, The 11-aminoacid long Tarsal-less peptides trigger a cell signal in Drosophila leg development, *Developmental Biology*, 324(2), 192–201.

**Qian, W., Choi, S., Gibson, G.A., Watkins, S.C., Bakkenist, C.J. & Van Houten, B.,** 2012, Mitochondrial hyperfusion induced by loss of the fission protein Drp1 causes ATM-dependent G2/M arrest and aneuploidy through DNA replication stress, *Journal of Cell Science*, 125(Pt 23), 5745–5757.

**Qurashi, A., Sahin, H.B., Carrera, P., Gautreau, A., Schenck, A. & Giangrande, A.,** 2007, HSPC300 and its role in neuronal connectivity, *Neural Development*, 2, 18.

**Rambold, A.S., Cohen, S. & Lippincott-Schwartz, J.,** 2015, Fatty acid trafficking in starved cells: regulation by lipid droplet lipolysis, autophagy, and mitochondrial fusion dynamics, *Developmental Cell*, 32(6), 678–692.

**Rambold, A.S., Kostelecky, B., Elia, N. & Lippincott-Schwartz, J.,** 2011, Tubular network formation protects mitochondria from autophagosomal degradation during nutrient starvation, *Proceedings of the National Academy of Sciences of the United States of America*, 108(25), 10190–10195.

**Rampersad, S.N.,** 2012, Multiple applications of Alamar Blue as an indicator of metabolic function and cellular health in cell viability bioassays, *Sensors*, 12(9), 12347–12360.

**Redpath, C.J., Bou Khalil, M., Drozdzal, G., Radisic, M. & McBride, H.M.,** 2013, Mitochondrial hyperfusion during oxidative stress is coupled to a dysregulation in calcium handling within a C2C12 cell model, *PloS one*, 8(7), e69165.

**Rehorn, K.-P., Thelen, H., Michelson, A.M. & Reuter, R.,** 1996, A molecular aspect of hematopoiesis and endoderm development common to vertebrates and Drosophila, *Development*, 122(12), 4023–4031.

**Rera, M., Bahadorani, S., Cho, J., Koehler, C.L., Ulgherait, M., Hur, J.H., Ansari, W.S., Lo, T., Jones, D.L. & Walker, D.W.,** 2011, Modulation of longevity and tissue homeostasis by the Drosophila PGC-1 homolog, *Cell Metabolism*, 14(5), 623–634.

**Roy, S., Ernst, J., Kharchenko, P.V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L., Lin, M.F., Washietl, S., Arshinoff, B.I., Ay, F., Meyer, P.E., and consortium,** 2010, Identification of functional elements and regulatory circuits by Drosophila modENCODE, *Science*, 330(6012), 1787–1797.

**Rusten, T.E., Lindmo, K., Juhász, G., Sass, M., Seglen, P.O., Brech, A. & Stenmark, H.,** 2004, Programmed autophagy in the Drosophila fat body is induced by ecdysone through regulation of the PI3K pathway, *Developmental Cell*, 7(2), 179–192.

**Sampson, C.J. & Williams, M.J.,** 2012, Real-time analysis of Drosophila post-embryonic haemocyte behaviour, *PloS one*, 7(1), e28783.

**Sampson, C.J., Amin, U. & Couso, J.P.,** 2013, Activation of Drosophila hemocyte motility by the ecdysone hormone, *Biology Open*, 2(12), 1412–1420.

**Sardiello, M.,** 2003, MitoDrome: a database of Drosophila melanogaster nuclear genes encoding proteins targeted to the mitochondrion, *Nucleic Acids Research*, 31(1), 322–324.

**Schatz, G. & Dobberstein, B.,** 1996, Common principles of protein translocation across membranes, *Science*, 271(5255), 1519–1526.

**Schmidt, O., Pfanner, N. & Meisinger, C.,** 2010, Mitochondrial protein import: from proteomics to functional mechanisms, *Nature Reviews Molecular Cell Biology*, 11(9), 655–667.

**Schneider, I.,** 1972, Cell lines derived from late embryonic stages of Drosophila melanogaster, *Journal of Embryology and Experimental Morphology*, 27(2), 353–365.

**Schoenberg, D.R. & Maquat, L.E.,** 2012, Regulation of cytoplasmic mRNA decay, *Nature Reviews Genetics*, 13(4), 246–259.

**Sharma, A.K. & FitzGerald, D.,** 2010, Pseudomonas exotoxin kills Drosophila S2 cells via apoptosis, *Toxicon*, 56(6), 1025–1034.

**Shim, J., Lee, S.M., Lee, M.S., Yoon, J., Kweon, H.S. & Kim, Y.J.,** 2010, Rab35 mediates transport of Cdc42 and Rac1 to the plasma membrane during phagocytosis, *Molecular Cell Biology*, 30(6), 1421–1433.

**Slavoff, S.A., Heo, J., Budnik, B.A., Hanakahi, L.A. & Saghatelian, A.,** 2014, A human short open reading frame (sORF)-encoded polypeptide that stimulates DNA end joining, *The Journal of Biological Chemistry*, 289(16), 10950–10957.

**Slavoff, S.A., Mitchell, A.J., Schwaid, A.G., Cabili, M.N., Ma, J., Levin, J.Z., Karger, A.D., Budnik, B.A., Rinn, J.L. & Saghatelian, A.,** 2012, Peptidomic discovery of short open reading frame–encoded peptides in human cells, *Nature Chemical Biology*, 9(1), 59–64.

**Smith, J.E., Alvarez-Dominguez, J.R., Kline, N., Huynh, N.J., Geisler, S., Hu, W., Coller, J. & Baker, K.E.,** 2014, Translation of small open reading frames within unannotated RNA transcripts in Saccharomyces cerevisiae, *Cell Reports*, 7(6), 1858–1866.

**Stein, L.D.,** 2004, Human genome: end of the beginning, *Nature*, 431(7011), 915-916.

**Stein, M., Dong, J. & Wandinger-Ness, A.,** 2003, Rab proteins and endocytic trafficking: potential targets for therapeutic intervention, *Advanced Drug Delivery Reviews*, 55(11), 1421–1437.

**Steller, H.,** 2008, Regulation of apoptosis in Drosophila, *Cell Death & Differentiation*, 15(7), 1132–1138.

**Stornaiuolo, M., Lotti, L.V., Borgese, N., Torrisi, M.R., Mottola, G., Martire, G. & Bonatti, S.,** 2003, KDEL and KKXX retrieval signals appended to the same reporter protein determine different trafficking between endoplasmic reticulum, intermediate compartment, and Golgi complex, *Molecular Biology of the Cell*, 14(3), 889–902.

**Su, B., Wang, X., Zheng, L., Perry, G., Smith, M.A. & Zhu, X.,** 2010, Abnormal mitochondrial dynamics and neurodegenerative diseases, *Biochimica et biophysica acta*, 1802(1), 135-142.

**Su, T.T., Parry, D.H., Donahoe, B., Chien, C.T., O'Farrell, P.H. & Purdy, A.,** 2001, Cell cycle roles for two 14-3-3 proteins during Drosophila development, *Journal of Cell Science*, 114(Pt 19), 3445–3454.

**Suen, D.F., Norris, K.L. & Youle, R.J.,** 2008, Mitochondrial dynamics and apoptosis, *Genes & Development*, 22(12), 1577–1590.

**Susin, S.A., Zamzami, N., Castedo, M., Hirsch, T., Marchetti, P., Macho, A., Daugas, E., Geuskens, M. & Kroemer, G.,** 1996, Bcl-2 inhibits the mitochondrial release of an apoptogenic protease, *The Journal of Experimental Medicine*, 184(4), 1331–1341.

**Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A. & Mello, C.C.,** 1999, The rde-1 gene, RNA interference, and transposon silencing in C. elegans, *Cell*, 99(2), 123–132.

**Tait, S.W. & Green, D.R.,** 2012, Mitochondria and cell signalling, *Journal of Cell Science*, 125(Pt 4), 807–815.

**Takahashi, K., Maruyama, M., Tokuzawa, Y., Murakami, M., Oda, Y., Yoshikane, N., Makabe, K.W., Ichisaka, T. & Yamanaka, S.,** 2005, Evolutionarily conserved non-AUG translation initiation in NAT1/p97/DAP5 (EIF4G2), *Genomics*, 85(3), 360–371.

**Tanner, E.A., Blute, T.A., Brachmann, C.B. & McCall, K.,** 2011, Bcl-2 proteins and autophagy regulate mitochondrial dynamics during programmed cell death in the Drosophila ovary, *Development*, 138(2), 327-338.

**Tepass, U., Fessler, L.I., Aziz, A. & Hartenstein, V.,** 1994, Embryonic origin of hemocytes and their relationship to cell death in Drosophila, *Development*, 120(7), 1829–1837.

**Thomenius, M., Freel, C.D., Horn, S., Krieser, R., Abdelwahid, E., Cannon, R., Balasundaram, S., White, K. & Kornbluth, S.,** 2011, Mitochondrial fusion is regulated by Reaper to modulate Drosophila programmed cell death, *Cell Death & Differentiation*, 18(10), 1640–1650.

**Thummel, C.S.,** 1996, Flies on steroids--Drosophila metamorphosis and the mechanisms of steroid hormone action, *Trends in Genetics*, 12(8), 306–310.

**Tiefenböck, S.K., Baltzer, C., Egli, N.A. & Frei, C.,** 2010, The Drosophila PGC-1 homologue Spargel coordinates mitochondrial activity to insulin signalling, *The EMBO Journal*, 29(1), 171–183.

**Toggas, S.M., Krady, J.K. & Billingsley, M.L.,** 1992, Molecular neurotoxicology of trimethyltin: identification of stannin, a novel protein expressed in trimethyltin-sensitive cells, *Molecular Pharmacology*, 42(1), 44–56.

**Tondera, D., Grandemange, S., Jourdain, A., Karbowski, M., Mattenberger, Y., Herzig, S., Da Cruz, S., Clerc, P., Raschke, I., Merkwirth, C., Ehses, S., Krause, F., Chan, D.C., Alexander, C., Bauer, C., Youle, R., Langer, T. & Martinou, J.C.,** 2009, SLP-2 is required for stress-induced mitochondrial hyperfusion, *The EMBO Journal*, 28(11), 1589–1600.

**Torres, T.T., Dolezal, M., Schlötterer, C. & Ottenwälder, B.,** 2009, Expression profiling of Drosophila mitochondrial genes via deep mRNA sequencing, *Nucleic Acids Research*, 37(22), 7509–7518.

**Truman, J.W., Talbot, W.S., Fahrbach, S.E. & Hogness, D.S.,** 1994, Ecdysone receptor expression in the CNS correlates with stage-specific responses to ecdysteroids during Drosophila and Manduca development, *Development*, 120(1), 219–234.

**Tycowski, K.T. & Steitz, J.A.,** 2001, Non-coding snoRNA host genes in Drosophila: expression strategies for modification guide snoRNAs, *European Journal of Cell Biology*, 80(2), 119–125.

**Tzou, P., De Gregorio, E. & Lemaitre, B.,** 2002, How Drosophila combats microbial infection: a model to study innate immunity and host-pathogen interactions,*Current Opion in Microbiology*, 5(1), 102–110.

**Ulvila, J., Vanha-Aho, L.M. & Rämet, M.,** 2011a, Drosophila phagocytosis - still many unknowns under the surface, *APMIS*, 119(10), 651–662.

**Ulvila, J., Vanha-aho, L.M., Kleino, A., Vähä-Mäkilä, M., Vuoksio, M., Eskelinen, S., Hultmark, D., Kocks, C., Hallman, M., Parikka, M. & Rämet, M.,** 2011b, Cofilin regulator 14-3-3eta is an evolutionarily conserved protein required for phagocytosis and microbial resistance, *Journal of Leukocyte Biology*, 89(5), 649–659.

**Vance, J.E., Stone, S.J. & Faust, J.R.,** 1997, Abnormalities in mitochondria-associated membranes and phospholipid biosynthetic enzymes in the mnd/mnd mouse model of neuronal ceroid lipofuscinosis, *Biochimica et Biophysica Acta*, 1344(3), 286–299.

**van den Eijnde, S.M., Boshart, L., Baehrecke, E.H., De Zeeuw, C.I., Reutelingsperger, C.P.M. & Vermeij-Keers, C.,** 1998, Cell surface exposure of phosphatidylserine during apoptosis is phylogenetically conserved, *Apoptosis*, 3(1), 9–16.

**Vanderperre, B., Lucier, J.F., Bissonnette, C., Motard, J., Tremblay, G., Vanderperre, S., Wisztorski, M., Salzet, M., Boisvert, F.M. & Roucou, X.,** 2013, Direct detection of alternative open reading frames translation products in human significantly expands the proteome, *PloS one*, 8(8), e70698.

**Vaucheret, H., Béclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J.-B., Mourrain, P., Palauqui, J.-C. & Vernhettes, S.,** 1998, Transgene-induced gene silencing in plants, *The Plant Journal*, 16(6), 651–659.

**Velichkova, M., Juan, J., Kadandale, P., Jean, S., Ribeiro, I., Raman, V., Stefan, C. & Kiger, A.A.,** 2010, Drosophila Mtm and class II PI3K coregulate a PI(3)P pool with cortical and endolysosomal functions, *The Journal of Cell Biology*, 190(3), 407–425.

**Vlisidou, I., Dowling, A.J., Evans, I.R. Waterfield, N. & Wood, W.,** (2009) Drosophila embryos as model systems for monitoring bacterial infection in real time, *PLoS Pathogens*, 5:e1000518.

**De Vos, K.,** 1998, The 55-kDa Tumour Necrosis Factor Receptor Induces Clustering of Mitochondria through Its Membrane-proximal Region, *The Journal of Biological Chemistry*, 273(16), 9673–9680.

**Wang, S.L., Hawkins, C.J., Yoo, S.J., Müller, H.-A.J. & Hay, B.A.,** 1999, The Drosophila caspase inhibitor DIAP1 is essential for cell survival and is negatively regulated by HID, *Cell*, 98(4), 453–463.

**Wang, Y., Nartiss, Y., Steipe, B., McQuibban, G.A. & Kim, P.K.,** 2012, ROS-induced mitochondrial depolarization initiates PARK2/PARKIN-dependent mitochondrial degradation by autophagy, *Autophagy*, 8(10), 1462–1476.

**Watts, C.,** 2012, The endosome-lysosome pathway and information generation in the immune system, *Biochimica et Biophysica Acta*, 1824(1), 14–21.

**Westermann, B.,** 2010, Mitochondrial fusion and fission in cell life and death, *Nature Reviews Molecular Cell Biology*, 11(12), 872–884.

**Westrate, L.M., Sayfie, A.D., Burgenske, D.M. & MacKeigan, J.P.,** 2014, Persistent mitochondrial hyperfusion promotes G2/M accumulation and caspase-dependent cell death, *PloS one*, 9(3), e91911.

**White, K., Grether, M.E., Abrams, J.M., Young, L., Farrell, K. & Steller, H.,** 1994, Genetic control of programmed cell death in Drosophila, *Science*, 264(5159), 677-683.

**Wood, W. & Jacinto, A.,** 2007, Drosophila melanogaster embryonic haemocytes: masters of multitasking, *Nature Reviews Molecular Cell Biology*, 8(7), 542–551.

**Yamanaka, N., Rewitz, K.F. & O'Connor, M.B.,** 2013, Ecdysone control of developmental transitions: lessons from Drosophila research, *Annual Review of Entomology*, 58, 497–516.

**Young, M.J., Bay, D.C., Hausner, G. & Court, D.A.,** 2007, The evolutionary history of mitochondrial porins, *BMC Evolutionary Biology*, 7, 31.

**Zambon, R.A., Vakharia, V.N. & Wu, L.P.,** 2006, RNAi is an antiviral immune response against a dsRNA virus in Drosophila melanogaster, *Cell Microbiology*, 8(5), 880–889.

**Zettervall, C.-J., Anderl, I., Williams, M.J., Palmer, R., Kurucz, E., Ando, I. & Hultmark, D.,** 2004, A directed screen for genes involved in Drosophila blood cell activation, *Proceedings of the National Academy of Sciences of the United States of America*, 101(39), 14192–14197.

**Zhang, Z. & Dietrich, F.,** 2005, Identification and characterization of upstream open reading frames (uORF) in the 5′ untranslated regions (UTR) of genes in Saccharomyces cerevisiae, *Current Genetics*, 48(2), 77–87.

**Zhao, J., Liu, T., Jin, S.B., Tomilin, N., Castro, J., Shupliakov, O., Lendahl, U. & Nistér, M.,** 2009, The novel conserved mitochondrial inner-membrane protein MTGM regulates mitochondrial morphology and cell proliferation, *Journal of Cell Science*, 122(Pt 13), 2252–2262.

**Zhou, Q., Kee, Y.S., Poirier, C.C., Jelinek, C., Osborne, J., Divi, S., Surcel, A., Will, M.E., Eggert, U.S., Müller-Taubenberger, A., Iglesias, P.A., Cotter, R.J. & Robinson, D.N.,** 2010, 14-3-3 coordinates microtubules, Rac, and myosin II to control cell mechanics and cytokinesis, *Current Biolgy*, 20(21), 1881–1889.

**Zimmermann, K.C., Ricci, J.E., Droin, N.M. & Green, D.R.,** 2002, The role of ARK in stress-induced apoptosis in Drosophila cells, *The Journal of Cell Biology*, 156(6), 1077–1087.

# Appendix 1

RT-PCR primers

DIAP1 forward primer: TGTGTGAAGTACATGCCGGT
DIAP1 reverse primer: TTCAGAGGAAAGGAGCCAGA
Product: 288bp
TGTGTGAAGTACATGCCGGTttgggccaattggcgtttatccagccagtccagcggccagtcggtgaaggt
ctttaatcgcgtctcctcgcggtttaaatcgttcattctggttttgtttatattatttttgaatagctgggtcgcgttcgtgttgttatccac
ctgatcaaaagcgataggtccataagacggaagatcagctacaacagatgccattttgtgttctggttgttagctcctttgtttgcc
tgactcttaatttcttcTCTGGCTCCTTTCCTCTGAA

Cdc2 forward primer: ACCAGATATCGACGGGACAG
Cdc2 reverse primer: GCTGCCAGTTGATAAGCACA
Product: 287bp
ACCAGATATCGACGGGACAGgaataccggggtgaacccagtagcacctccggcgctctgtaccacaaggt
aacaatctcgtgcgtataaatgcgcaccggaatgccaaaggatcggccaagtccaaagtcggcgacttttatgaggccac
tcttgtcgattagtaagttctgcggcttaagatcacggtgaagtactcgccgacgatggcagaaaagaatggcgctagttatttg
gtacaaatagctacggaccaattcactctccaTGTGCTTATCAACTGGCAGC

Rp49 forward primer: CCAGTCGGATCGATATGCTAA
Rp49 reverse primer: GAGGTCCTGCTCATGCAGA
Product:268bp

CCAGTCGGATCGATATGCTAAGCTGTCGGTGAGTGCCAACGAGGATTGTGCC
AAATTGTACCCGTGTTTAATCAACATGTCTCCTTGCAGCACAAATGGCGCAAG
CCCAAGGGTATCGACAACAGAGTGCGTCGCCGCTTCAAGGGACAGTATCTGA
TGCCCAACATCGGTTACGGATCGAACAAGCGCACCCGCCACATGCTGCCCAC
CGGATTCAAGAAGTTCCTGGTGCACAACGTGCGCGAGCTGGAGGTCCTGCT
CATGCAGA

# Appendix 2

dsRNA primers

(T7 polymerase priming sequence underlined)

Diap1Fw TAATACGACTCACTATAGGGAGCCTGTTGAACCACGATTTT

Diap1Rv TAATACGACTCACTATAGGGATCTCATGGATTGGAACGACA

Drp1Fw TAATACGACTCACTATAGGGATTGAGAACCCGAACTCCATC

Drp1Rv TAATACGACTCACTATAGGGATAAGATCTGGCAGGCAGTCC

Pi3K68DFw TAATACGACTCACTATAGGGAGGCAAGCGCATATCGACTACGACA

Pi3K68DRvTAATACGACTCACTATAGGGAGGGTGGAAAGCCGTATTGAGGAGG

14-3-3Fw TAATACGACTCACTATAGGGAGGGGGACAACCTGACTCTCTGGAC

14-3-3Rv TAATACGACTCACTATAGGGAGGACTGTGAAGCATTCTGCCCTCA

CG33714Fw TAATACGACTCACTATAGGGAGGACTACAAGCAGGCCACGAAT

CG33714Rv TAATACGACTCACTATAGGGAGGCTCCGCACCAGGAACTGTAT

CG32276Fw TAATACGACTCACTATAGGGAGGAGGCTTCACCTCGCATAAGA

CG32276Rv TAATACGACTCACTATAGGGAGGATTTAAAACTGCGCCGAGAA

CG14104Fw TAATACGACTCACTATAGGGAGGAAGAAAAAGTCGCGCAAAAC

CG14104Rv TAATACGACTCACTATAGGGAGGACAAACTTTATTCGTCGCGG

CG14036Fw TAATACGACTCACTATAGGGAGGTTTTGACAAACGGCTCTTGA

CG14036Rv TAATACGACTCACTATAGGGAGGGGATGCCTGCTTGTCTTAGC

CG6770Fw TAATACGACTCACTATAGGGAGGCAAAAAGCAGACACCAACCA

CG6770Rv TAATACGACTCACTATAGGGAGGAGCTTGGTCAGAATCTTGCG

CG34242Fw TAATACGACTCACTATAGGGAGGTGGTTGGAGATTCGTTGGAT

CG34242Rv TAATACGACTCACTATAGGGAGGGGAGTTCATGCTGTCAGGGT

CR43250Fw TAATACGACTCACTATAGGGAGGGAGCGGCTAAAAATATCCGA

CR43250Fw TAATACGACTCACTATAGGGAGGGGGTTCTGTGAAAAAGCGAA

CG9878Fw TAATACGACTCACTATAGGGAGGACGCATTTGTGCTGTCAAAG

CG9878Rv TAATACGACTCACTATAGGGAGGTTTCGATCTCCATTTCCTGC

CG42371Fw TAATACGACTCACTATAGGGAGGAAAATTGGTGAAATCGTCGG

CG42371Fw TAATACGACTCACTATAGGGAGGAAGCCAATCAGCCACAAAAC

CG32736Fw TAATACGACTCACTATAGGGAGGATGCCCGCCGGAGTTT

CG32736Rv TAATACGACTCACTATAGGGAGGCTTTGCGGTGGGATCCAC

CG17278Fw TAATACGACTCACTATAGGGAGGACTCCTTGTCGCTGATTTTTT

CG17278Rv TAATACGACTCACTATAGGGAGGGACACGCAGGAGTACCGATC

HemotinFw TAATACGACTCACTATAGGGAGGTTCAATCGGAAATCAATCCCTTA

HemotinRv TAATACGACTCACTATAGGGAGGGTCATCGTGATGCACTGTCCAA

CG11686Fw TAATACGACTCACTATAGGGACGCTCTGAACAAACAACACC

CG11686Rv TAATACGACTCACTATAGGGAGCAAATGCGTAGACCCAGAT

CG34293Fw TAATACGACTCACTATAGGGACTTTTGTGCCACATAGGCTG

CG34293Rv TAATACGACTCACTATAGGGACCCTGGGAAAAGAAGCTGAT

CG12012Fw TAATACGACTCACTATAGGGATGCATCAGGGAGCACCTAGT

CG12012Rv TAATACGACTCACTATAGGGATAGAAAACCGTTCCGCAGTT

CG32448Fw:TAATACGACTCACTATAGGGATGCTCCCAATTCGACTTCTT

CG32448Rv TAATACGACTCACTATAGGGACAAGATTATCATGGGCCTGG

CG18343Fw TAATACGACTCACTATAGGGAAGGAAATGTCCCATTTGTCG

CG18343RvTAATACGACTCACTATAGGGAACAGATGTGATGTCGACGGA

CG31360Fw TAATACGACTCACTATAGGGACGCCTTATTTTCCAAATCCA

CG31360Rv TAATACGACTCACTATAGGGATCCAGAGGTTTCGCTTGTCT

CG32267Fw TAATACGACTCACTATAGGGAAGGATTTGGGTACATTGCCA

CG32267RvTAATACGACTCACTATAGGGATTTTTGGGCTGAAAGGTACG

CG9669Fw TAATACGACTCACTATAGGGATATGTTAACGCCCCCc

CG9669Rv TAATACGACTCACTATAGGGACTACGTATCGCCCGTGAAC

CG13784FwTAATACGACTCACTATAGGGACCAGGCACACGTACAGAGAG

CG13784Rv TAATACGACTCACTATAGGGAAATGATGAGATGGGCCTGAC

CG17680Fw TAATACGACTCACTATAGGGAAATATATGGCCGTGAGCTGG

CG17680Rv TAATACGACTCACTATAGGGACCACATGTCCAGCGTGTACT

CG4101Fw TAATACGACTCACTATAGGGAATGTGGGTGCTCTCTTCCAG

CG4101Rv TAATACGACTCACTATAGGGACGTCGCTTCTTCTTCTCCAG

CG18622Fw TAATACGACTCACTATAGGGATCTCCATCAAATACCAGGGC

CG18622Rv TAATACGACTCACTATAGGGAAAGGACTCTCCGGAGTCGAA

# Appendix 3 - Papers published